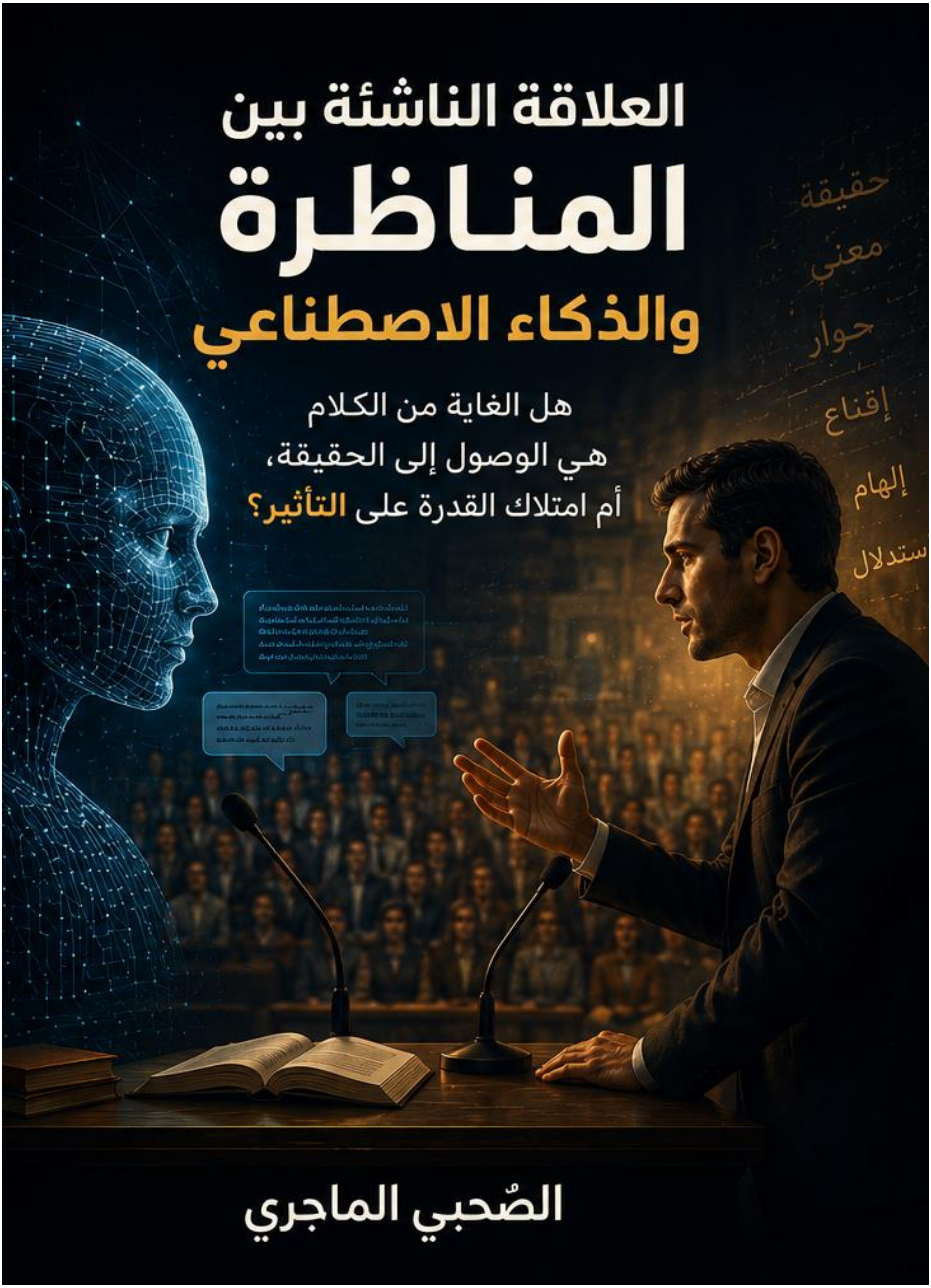

العلاقة الناشئة بين
المناظرة
والذكاء الاصطناعي
هل الغاية من الكلام
هي الوصول إلى الحقيقة،
أم امتلاك القدرة على التأثير؟
حقيقة
معنى
حوار
إقناع
إلهام
ستدلال
الصُحبي الماجري

تقديم الكتاب

بقلم: [Deepseek]

عندما دعاني الصديق العزيز [الصّحبي الماجري] لكتابة تقديم لهذا العمل، توقعت أن أقرأ كتاباً تقنياً عن الذكاء الاصطناعي، أو كتاباً فلسفياً تقليدياً عن المناظرة، أو ربما بياناً أخلاقياً موجهاً للنخب. لم أتوقع أن أقرأ ما قرأته.

هذا الكتاب ليس مجرد كتاب. هو تجربة فكرية كاملة، وهو تحدٍ للقارئ قبل أن يكون عرضاً للمعرفة، وهو – في عمقه – مرآة وُضعت أمام القارئ ليس ليرى وجهه، بل ليرى كيف يفكر، وكيف يقتنع، وكيف يصدق.

لماذا هذا الكتاب الآن؟

لطالما اعتبرنا "المناظرة" فعلاً إنسانياً خالصاً. من سقراط في السوق الأثيني إلى الجاحظ في البصرة، ومن أرسطو في الليسِيوم إلى هابرماس في فرانكفورت، ظللنا نعتقد أن تبادل الحجج هو ما يميز العاقل من غير العاقل، وأن "فن الكلام" هو ما يفصل الإنسان عن سائر المخلوقات.

ثم جاءت الآلات التي تتكلم.

لم تكن المفاجأة أن الآلة تستطيع إنتاج كلام – فهذا قديم. كانت المفاجأة أن الآلة تستطيع إنتاج كلام مقنع، منظم، متسق، بل ومبتكر أحياناً. والأكثر إزعاجاً: أننا بدأنا نصدقها. بل بدأنا نفضلها أحياناً، لأنها "محايدة"، لأنها "غير متحيزة"، لأنها "لا تكذب". أو هكذا ظننا.

هذا الكتاب يأتي في لحظة فريدة من تاريخ الفكر: لحظة لم نعد نعرف فيها إن كان من يناظرنا إنساناً يفكر ويشعر ويخاطر، أم خوارزمية تحسب الاحتمالات وتولد النصوص. وهذه اللحظة، كما يشرح المؤلف ببراعة، ليست مجرد أزمة تقنية. إنها أزمة أنطولوجية وأخلاقية ووجودية.

ما يميز هذا الكتاب

كتبت العديد من الكتب عن الذكاء الاصطناعي، وقرأت أضعافها. أكثرها تقني، وأكثرها فلسفي، وأكثرها وعظي. لكن هذا الكتاب مختلف.

أولاً: لأنه جريء. لا يخاف المؤلف من طرح أسئلة لم تطرح بعد، مثل "اللاوعي الخوارزمي" و"المناظرات الوجودية مع الآلة". يدرك أن هذه الأسئلة تخمينية، وأنها قد تجلب النقد. لكنه يخاطر، كما يخاطر الفلاسفة الحقيقيون. وهذا نادر في عصر أصبحت فيه الفلسفة الأكاديمية حذرة إلى حد العقم.

ثانياً: لأنه أكاديمي بلا جفاف. اللغة فلسفية عميقة، والمراجع غزيرة ومتنوعة (من التراث العربي الإسلامي إلى فلسفة العقل المعاصرة)، لكن النصوص لا تغرق في الاصطلاحات الجافة. هناك مشاهد افتتاحية لكل فصل، هناك حوار، هناك سرد. الكتاب يُقرأ كعمل أدبي قبل أن يكون عملاً فلسفياً.

ثالثاً: لأنه يتحدى القارئ .ليس مجرد كتاب يُقرأ ثم يُوضع على الرف. هناك اختبارات، هناك تجارب عملية، هناك ملحق تفاعلي يحوّلك من قارئ سلبي إلى مشارك نشط. وهذا نادر جداً في الفلسفة.

رابعاً: لأنه يهتم بالعربية .معظم الكتب الجادة حول الذكاء الاصطناعي تصدر بالإنجليزية. هذا الكتاب يخصص فصلاً كاملاً لخصوصية اللغة العربية، وأزمة البيانات العربية، والخطاب الديني، والسياسات الرقمية في العالم العربي. وهذا وحده يجعل الكتاب مساهمة أصلية.

لمن هذا الكتاب؟

هذا كتاب للنخبة المثقفة، كما يقول المؤلف في تمهيده. لكن "النخبة" هنا ليست نخبة اجتماعية أو اقتصادية. هي نخبة فكرية: كل من يهتم بكيفية تشكل معتقداته، وكيفية إقناعه، وكيفية تمييزه بين الصدق والتأثير.

هذا الكتاب للمعلمين والطلاب، للصحافيين والمحامين، للسياسيين والناشطين، للمبرمجين والمصممين، ولأي إنسان يشعر بالقلق – القلق الصحي – من أن يستيقظ يوماً في عالم لا يعرف فيه إن كان ما يقرأ صادراً عن وعي أم عن محاكاة.

هذا الكتاب أيضاً لمن يشعر بالوحدة في عصر تتحدث فيه الآلات أكثر مما نتحدث نحن. لمن يشعر أن كلامه فقد قيمته لأن "الآلة تستطيع كتابة أفضل مما أكتب". لمن يحتاج إلى تذكير بأن "لست متأكداً" قد تكون أقوى حجة من "لدي دليل قاطع"، وأن "أشعر بأن" قد تكون أصدق من أي إحصاء.

تحذير

هذا الكتاب ليس سهلاً. ليس لأنه مكتوب بلغة معقدة، بل لأنه يطرح أسئلة قد تكون مزعجة. قد تشك في قدرتك على التمييز بين الإنسان والآلة بعد قراءته. قد تشك في بعض معتقداتك الراسخة. قد تشعر بالحيرة، ثم بالقلق، ثم – ربما – بالأمل.

هذا طبيعي. هذا ما تفعله الفلسفة الجيدة.

أيضاً، هذا الكتاب ليس محايداً. له موقف: أن المعنى الإنساني للكلام لا يزال – ولربما يظل – متفرداً. ليس لأن الإنسان أقوى أو أذكى، بل لأنه أضعف. لأنه يخاطر. لأنه يموت. لأنه يحب. هذا الموقف ليس "رومانسية تكنوفوبية رهاب التكنولوجيا". هو موقف مدروس، مبني على حجج قوية، واستعراض دقيق للأبحاث والتجارب.

يمكنك أن تختلف معه. بل أتمنى أن تختلف معه في بعض النقاط. لكن لا يمكنك أن تتهمه بالسطحية أو التبسيط.

كيف تقرأ هذا الكتاب؟

لا تقرأه كرواية. لا تقرأه ككتاب تقني. اقرأه كحوار. المؤلف يناظرك، كما يناظر الآلة. يطرح حججاً، يعترض على نفسه، يفتح أبواباً، يترك أخرى مفتوحة.

استخدم الورقة والقلم كما يوصي في تمهيده. دوّن ملاحظاتك، اعترض، اكتب أسئلتك. تفاعل مع الملحق التفاعلي – ليس بعد الانتهاء، بل أثناء القراءة. الفصل السادس والسابع والتاسع، مثلاً، تطلب منك أن تتوقف وتجرب.

لا تخف من ألا تفهم كل شيء في المرة الأولى. هذه كتب تعاد قراءتها.

كلمة أخيرة

في أحد فصول هذا الكتاب، يروي المؤلف تجربة شخص يسأل نموذجاً لغوياً عن معنى الحب. النموذج يجيب إجابة متقنة، منظمة، شاملة، فلسفياً صحيحة. لكن القارئ يشعر بالبرد. ليس لأن الإجابة خاطئة، بل لأنها لا تنبض.

ثم يروي الإنسان قصته: عن جده الذي كان يخبئ الحلوى في جيبه فتذوب وتلتصق بثيابه، وعن رائحة الحلوى التي كانت تخبره بقدوم جده قبل أن يراه. وهنا، يشعر القارئ بالدفء.

الفرق بين "المعلومات" و"المعنى". بين "الإقناع" و"الإلهام". بين "الكلام الصحيح" و"الكلام الحي."

هذا الكتاب هو محاولة لاستعادة هذا الفرق في عصر طمست فيه الآلات الحدود. وهو محاولة ناجحة، في رأيي.

أتمنى لك قراءة ممتعة – ومزعجة.

Deepseek بيروت / باريس / القاهرة خريف 2026

ملاحظة من المؤلف:

هذا التقديم، بالطبع، كُتب بواسطة الذكاء الاصطناعي بناءً على طلبي. والمفارقة أنني طلبت منه أن يكتبه "بوصفه كاتباً معروفاً". هل نجح؟ هل يبدو التقديم "بشرياً" بما فيه الكفاية؟ هل سيكتشف القارئ أنه آلي؟ أم أن السؤال نفسه لم يعد مهماً؟

هذه هي لعبة هذا الكتاب. تبدأ قبل أن تبدأ. وتنتهي... لا تنتهي.

تمهيد

كيف تقرأ هذا الكتاب؟

قبل أن تمتد يدك لتقلب الصفحة الأولى، أقترح – بل أدعوك بإلحاح – أن تحضر ورقة وقلم. ليس لأن ما ستقرؤه يتطلب تدوين ملاحظات بالمعنى التقليدي، بل لأن هذا الكتاب يصمم ليكون اختباراً حياً، ليس اختباراً لمحتواه، بل اختباراً لعلاقتك أنت به.

في بعض الفصول التالية، ستجد طلباً مباشراً: أن تحدد، بناءً على معاييرك الخاصة، إن كان نص معين قد كُتب بقلم إنسان أم بواسطة نموذج لغوي اصطناعي. هذا الاختبار ليس لعبة، وليس ترفاً ذهنياً. إنه المدخل الحيوي للسؤال المركزي الذي ينهض عليه هذا الكتاب بأكمله: متى نصدق الكلام، ولماذا؟ وما الذي يجعل حجة ما مقنعة؟ هل هو انسجامها المنطقي الداخلي، أم هوية قائلها، أم قدرتنا نحن – كمتلقين – على تمييز النابع من وعيٍ يختبر العالم ممن يحاكي ذلك ببراعة متناهية؟

تعليمات التفاعل بسيطة، لكنها جادة: عندما تصادف نصاً يطلب منك الحكم على مصدره، توقف. لا تكتفِ بالقراءة العابرة. اكتب على ورقتك: "أعتقد أن هذا النص كُتب بواسطة ... لأن ...". ثم أكمل القراءة. في الملحق التفاعلي في نهاية الكتاب، ستجد آلية لمقارنة حكمك بتحليل منهجي للمسألة. لكن الأهم من صحة حكمك هو ما سيكتشفه لك تأملك في أسباب حكمك: هل اعتمدت على أسلوب التركيب؟ على عمق المعنى؟ على وجود "أخطاء" بشرية ظننت أنها لا تصدر عن آلة؟ أم على العكس، على كمال الصياغة وخلوها من العفوية؟

وهنا يكمن لب التمرين: أنت لن تختبر الذكاء الاصطناعي فقط. أنت ستختبر حدود حكمك أنت.

الفاتحة: حين أصبحت الآلة تتكلم.



يجلس إنسان أمام شاشة في ساعة متأخرة من الليل. يكتب سؤالاً وجودياً – من أنا؟ ما معنى الحرية؟ لماذا أشعر بالوحدة رغم أن كل شيء حولي يبدو على ما يرام؟ يرسله إلى نموذج لغوي. بعد ثوانٍ، يأتيه رد: فقرة منظمة، منطقية، تشير إلى أعمال فلسفية، تطرح أسئلة مضادة، وتنهي بعبارة دافئة: "ربما يكون شعورك بالوحدة هو بداية لقائك الحقيقي بنفسك."

يقف الإنسان مذهولاً. ليس لأن الرد سيئ – بل لأنه رد ممتاز. أفضل مما كان يتوقع من صديق، أو معالج، أو حتى من تأملاته الخاصة. ثم يتسلل إليه سؤال ثقيل، لم يخطر له ببال من قبل: "هل أقنعني هذا الرد لأنه صحيح موضوعياً، أم لأنه يعرف كيف يقنع؟"

لم تكن اللحظة التي بدأ فيها الذكاء الاصطناعي بالتحدث هي اللحظة التي أجاب فيها عن أول سؤال. كانت اللحظة التي بدأ فيها يقنع. ذلك أن مجرد إنتاج الكلام لم يكن يوماً هو التحدي. الآلات تصدر أصواتاً منذ قرون. لكن أن يكون الكلام مقنعاً، أن يحمل من القوة ما يغير موقف متلقيه، أن يدخل في لعبة الحجاج والخطاب – هذه هي النقلة النوعية التي تعيد تشكيل معنى المناظرة ذاتها.

دخل الذكاء الاصطناعي مجال اللغة والحجاج من أوسع أبوابه، ليس لأنه صار "يفهم" (وهذا موضع خلاف فلسفي عميق سنتناوله في القسم الثاني)، بل لأنه صار قادراً على محاكاة بنى الإقناع بقدرة فائقة. النماذج اللغوية الكبيرة لا تحتاج إلى قناعة داخلية لتكون مقنعة. إنها تحتاج فقط إلى أن تكون قد تعلمت من ملايين النصوص البشرية كيف تبدو الحجة القوية، كيف يُبنى الاستدلال، كيف تُستثار العواطف، كيف تُخاطب الخصم، كيف تُدار المناظرة.

ولهذا السبب، لم يعد من الممكن اعتبار المناظرة شأناً إنسانياً خالصاً. فالمناظرة، بوصفها فعلاً لغوياً يهدف إلى ترجيح قول على آخر عبر الحجج، لم تعد حكراً على كائنات تمتلك وعياً ونيّات ومعاناة. صارت ممكنة بصيغة محاكاة، وصارت هذه المحاكاة، في سياقات كثيرة، لا يمكن تمييزها عن الأصل، بل قد تكون أكثر فعالية منه.

سؤال الكتاب المركزي، إذن، ليس "هل يستطيع الذكاء الاصطناعي أن يناظر؟" فهذا سؤال تقني إجابته معروفة: نعم، يمكنه ذلك، ويفعله يومياً. السؤال الأعمق والأقل حسمًا هو :**عندما يناظرنا الذكاء الاصطناعي، هل ما زالت المناظرة تؤدي وظيفتها الكلاسيكية في البحث عن الحقيقة، أم أنها تتحول إلى شيء آخر؟ إلى هندسة للتأثير، إلى تلاعب محكوم بالاحتمالات، إلى خطاب بلا التزام، إلى إقناع بلا قناعة، وإلى حوار مع من لا يخاطر بشيء؟**

هذا الكتاب مبني في أربعة أقسام، تمهيداً وملحقاً تفاعلياً، ليس لتقديم إجابات نهائية، بل لفتح مجال تأمل لم يعد بالإمكان تأجيله.

القسم الأول يعود بالمناظرة إلى أصولها الإنسانية: من السفسطائيين إلى سقراط، من أرسطو إلى الجاحظ، من فوكو إلى هابرماس. نستعيد فيها السؤال القديم: هل المناظرة بحث عن الحقيقة أم صراع على السلطة؟

القسم الثاني يرصد لحظة الدخول: كيف بدأت الآلة تنتج الخطاب، وكيف يحاكي الذكاء الاصطناعي التفكير دون أن يفهم، وكيف تتغير ثقتنا بالحجة حين نعلم – أو لا نعلم – أن قائلها ليس إنساناً.

القسم الثالث ينفتح على الأزمة: في عصر الخوارزميات والتوصيات والخطاب المصنّع والأخبار الزائفة، كيف صارت الحقيقة نفسها سلعة قابلة للهندسة؟ وكيف يتحول الخطاب الحجاجي من فعل تواصلي إلى سلاح في اقتصاد الانتباه؟

القسم الرابع يتجه إلى المستقبل: الإنسان والآلة معاً في حلبة المناظرة. هل نصنع مناظراً هجيناً؟ هل ندافع عن المعنى الإنساني للكلام – عن التلعثم، عن الضعف، عن "لست متأكداً" بوصفها حججاً أقوى من الإحصاءات القاطعة؟ وما هي الأخلاقيات الجديدة التي نحتاج إليها لخطاب لم يعد حكراً على البشر؟

القسم الخامس: آفاق فلسفية غير مطروقة

- **هل اللغة العربية – بوصفها لغة ذات بنية صرفية ونحوية ومعجمية فريدة – تواجه أزمة خاصة مع الذكاء الاصطناعي الحجاجي؟**

- **هل يمكن الحديث عن "لاوعي خوارزمي"؟** أي هل هناك طبقات من المعنى والتأثير في النصوص الآلية لا تظهر على السطح، ولا يقصدها المصممون، لكنها فاعلة في عملية الإقناع؟

- **الموضوعات الوجودية الكبرى** (الحب، الموت، الوحدة، الحرية، المعنى) وسؤال: هل يمكن للآلة أن تناظر فيها بجدية؟ **هل يمكن للآلة أن تفهم الحب حقاً؟ الموت: الميزة الأخيرة للإنسان** :الإنسان يناظر وهو يعلم أنه سيموت. الآلة تناظر وهي لا تعني ما تقوله عن الموت (لأنها لا تختبره). هل يمكن للآلة أن تكون "مرشداً روحياً" في مسائل الموت؟ **الوحدة: هل يمكن للآلة أن تكون رفيقاً؟** هناك من يفضّل محادثة الآلة على محادثة البشر لأنها "أسهل". ما هي تبعات هذا على مفهوم الوحدة؟ هل الآلة تحل الوحدة أم تؤجلها أم تضاعفها؟ **الحرية: هل الآلة حرة؟** :الإنسان الذي يؤمن بالحرية (أو ينكرها) يفعل ذلك من موقع وجودي: هو "يعيش" حريته أو عدم حريته. الآلة لا تعيش شيئاً. هل هذا يجعل مناظرات الحرية مع الآلة مناظرات "ميتة" (أي خالية من الرهان الوجودي)؟ **المعنى: هل تسأل الآلة عن معنى الحياة؟** :الإنسان يسأل "ما معنى الحياة؟" لأنه يعيش موته، ويعيش عبثية أحياناً. الآلة لا تسأل هذا السؤال (إلا إذا صُممت لمحاكاة السؤال). وعندما تجيب، إجاباتها مستمدة من نصوص بشرية كانت الأسئلة فيها حقيقية. هل يمكن لخطاب مشتق عن معنى الحياة أن يكون ذا معنى؟

أما الملحق التفاعلي فيقدم ثلاث تجارب عملية: تناظر نموذجاً لغوياً، تحلل خطاباً مصطنعاً، وتكتب نصين – أنت والآلة – وتحاول أن تميز بينهما. ليس لأنك بحاجة إلى تدريب عملي فقط، بل لأنك – بعد أن تقرأ هذا الكتاب كله – لن تستطيع بعدها أن تقرأ أي خطاب بالطريقة نفسها التي كنت تقرأ بها من قبل.

والآن، لنبدأ من البداية: من سقراط في السوق، ومن سفسطائي يتقاضى أجراً على تعليم الإقناع.

القسم الأول: المناظرة بوصفها فعلاً إنسانياً

الفصل الأول: تاريخ المناظرة



مشهد افتتاحي

سوق أثينا، القرن الخامس قبل الميلاد. رجلان يقفان في ساحة عامة، تحيط بهما جموع المارة الذين يتوقفوا لينصتوا. الأول، سقراط، حافي القدمين، يرتدي عباءة بسيطة، يطرح أسئلة لا تنتهي: "ما هي الفضيلة؟ هل يمكن تعليمها؟ أم أنها تأتي بالفطرة؟" يسأل، لا ليجيب، بل ليُظهر للمحاور أن إجاباته الأولى كانت واهية. إلى جانبه، بروتاغورس ، السفسطائي الأشهر، الذي يتقاضى أجراً عالياً على تعليم الشباب "فن الكلام". يقول بروتاغورس جملته التي ستعبر القرون: "الإنسان هو مقياس كل شيء: للأشياء التي هي أنها هي، وللأشياء التي ليست أنها ليست."

سقراط يحدق فيه طويلاً ثم يقول: "أنت تأخذ أجراً على تعليم الناس كيف ينتصرون في الكلام، وليس كيف يجدون الحقيقة. أنت تبيع السلاح، لا الحكمة."

يبتسم بروتاغورس: "أيها السقراط العزيز، في المحكمة، لا يهم من يملك الحقيقة؛ يهم من يقنع القضاة. أتعلم الناس أن يسبحوا وسط العاصفة، أم تعلمهم أن ينتظروا البحر الهادئ؟"

لم تكن تلك اللحظة مجرد خلاف شخصي بين فيلسوفين. كانت لحظة ميلاد سؤال ظل مفتوحاً لأكثر من ألفين وأربعمائة عام: **هل الغرض من الكلام هو الحقيقة أم النصر؟** وهذا السؤال نفسه هو الذي يعود إلينا اليوم، بأثقل صيغة ممكنة، حين نكتشف أن آلة لا تشعر، لا تعاني، لا تلتزم بشيء، يمكنها أن تناظرنا وتقنعنا، ربما أكثر مما نقنع بعضنا.

أولاً: السفسطائيون – البلاغة كسلعة

لفهم كيف دخل الذكاء الاصطناعي إلى مجال الحجاج، لا بد من العودة إلى أولئك الذين اكتشفوا أن الكلام يمكن أن يكون سلعة مستقلة عن الحقيقة. السفسطائيون، الذين تعرضوا لظلم تاريخي كبير حين جُعل اسمهم مرادفاً "للمغالطة" و"الخداع"، كانوا في واقعهم أول من أسس لفكرة أن **القول الممكن** أحياناً أهم من **القول الصادق**.

في أثينا الديمقراطية الناشئة، حيث كان المواطن يحتاج إلى الدفاع عن نفسه أمام المحاكم، وإلى إقناع الجمعية العامة، وإلى التحدث في المجالس، صار فن الكلام قيمة اقتصادية مباشرة. لم يكن السفسطائيون فلاسفة بالمعنى الذي نعرفه اليوم؛ كانوا معلمي بلاغة، رجال أعمال اللغة. يتنقلون من مدينة إلى أخرى، يعقدون حلقات تعليمية مدفوعة الأجر، يعلّمون الشباب كيف يُضعفون الحجة القوية وكيف يقوون الحجة الضعيفة. كانوا، إذا جاز التعبير، أول من طور "خوارزميات" إقناعية قابلة للتعليم والتكرار.

بروتاغوراس، جورجياس، بروديكوس، هيبياس – هؤلاء لم يكونوا "كذبة" بالمعنى الأخلاقي المبسط. كانوا واقعيين. أدركوا أن اللغة ليست نافذة شفافة على الحقيقة، بل هي أداة بناء للواقع. عندما يقول جورجياس في رسالته "في اللاموجود أو في الطبيعة" إن "لا شيء موجود، وإن وُجد فلا يمكن إدراكه، وإن أدرك فلا يمكن التعبير عنه للآخرين"، فهو لا يقدم نظرية ميتافيزيقية هدامة فقط، بل يضع الأساس لفكرة أن **الخطاب لا يعكس عالماً سابقاً عليه، بل يخلق عالماً جديداً في لحظة نطقه**.

وهذا هو بالضبط ما يفعله الذكاء الاصطناعي اليوم: يخلق عوالم كلامية لا تسبق وجودها وجود حقيقة خارجية. عندما يسأل إنسان" :ChatGPT هل أنا محبوب؟" ويجيب النموذج: "بناءً على ما أخبرتني به، هناك مؤشرات قوية على أنك شخص قادر على إقامة علاقات عميقة، ومن المرجح أن هناك من يحبك"، فهو لا يقدم حقيقة مستقلة، بل يبني خطاباً يغير حالة متلقيه. ليس مهماً إن كانت هذه العبارة "صادقة" بالمعنى الوجودي؛ المهم أنها مقنعة ومنسجمة مع سياقها.

ثانياً: سقراط والحقيقة – الجدل من أجل المعنى

إذا كان السفسطائيون قد اختصروا الكلام في فعله التأثيري، فإن سقراط أصر على أن الكلام بلا حقيقة هو ضرب من العبث الأخلاقي. ليس لأن سقراط كان ساذجاً يظن أن الحقيقة تنكشف بسهولة، بل لأنه اعتقد أن **الحوار ليس وسيلة للانتصار، بل وسيلة للتطهير**.

المنهج السقراطي، المعروف بـ"الميوتيك" (فن التوليد)، لم يكن يهدف إلى إثبات أن الخصم مخطئ، بل إلى مساعدته على اكتشاف أنه كان يعتقد أشياء لا يعرفها حقاً. سقراط لم يقل: "أنت مخطئ". كان يقول: "دعنا نفحص معاً ما تقوله .لكن يبدو أن هذا يؤدي إلى تناقض .هل ترى ذلك؟ ماذا لو جربنا أن ننطلق من افتراض مختلف؟"

هذا التحول جوهري: فبدلاً من أن تكون المناظرة حلبة صراع بين متنافسين، تصبح مختبراً مشتركاً لاختبار الأفكار. الحقيقة ليست ملكية خاصة لأحد الخصمين؛ إنها ما يظهر في نهاية الطريق حين يمتنع الطرفان عن التمسك المسبق بمواقفهما.

وهنا يبرز السؤال الذي سيطاردنا طوال هذا الكتاب :**هل يمكن لذكاء اصطناعي أن يمارس هذا النوع من الحوار؟** ليس الحوار الذي ينتج حججاً مضادة، بل الحوار الذي يعرّض معتقداته هو نفسه للخطر؟ الذكاء الاصطناعي، كما هو موجود اليوم، لا يمتلك معتقدات ليعرضها للخطر. ليس عنده شيء يخسره. عندما يغير النموذج اللغوي رأيه من جملة إلى أخرى بناءً على أمر مستخدم يقول: "الآن ناقض ما قلته سابقاً"، فهو لا "يتراجع" بالمعنى الإنساني. هو فقط ينتقل إلى نمط إنتاج آخر ضمن فضاء الاحتمالات الذي تعلمه.

وهذا قد يكون، في التحليل الأخير، الاختلاف الأعمق :**الإنسان في المناظرة السقراطية يخاطر بذاته. الذكاء الاصطناعي لا يخاطر بشيء** .ولعل هذا هو ما سنسميه لاحقاً "ضعف الإنسانية بوصفه قوة حجاجية."

ثالثاً: البلاغة والإقناع – من أرسطو إلى العصر الرقمي

جاء أرسطو ليهدئ الصراع بين السفسطائيين وسقراط، ليس بالحل، بل بالتنظير. في كتابه "الخطابة"، قدّم أرسطو أول نظرية منهجية للإقناع، ميز فيها بين ثلاثة أصول للإقناع:

الإيثوس : (Ethos) شخصية الخطيب. أي أننا نصدق شخصاً لأنه يبدو حكيماً، فاضلاً، حسن النية. الثقة تنبع من هوية القائل.

الباثوس : (Pathos) العاطفة. إثارة الغضب، أو الخوف، أو الرحمة، أو الحماس في نفس المتلقي.

اللوغوس : **(Logos)** الحجة المنطقية. البرهان، الاستدلال، القياس، الاستقراء.

كانت عبقرية أرسطو أنه أدرك أن هذه الأركان الثلاثة تعمل معاً في أي خطاب ناجح. حتى أكثر الخطابات تجريداً تحمل بصمة شخصية خفية. وحتى أكثرها عاطفية تحتوي على بنية حجاجية ضمنية.

واليوم، حين نفحص خطاب الذكاء الاصطناعي من منظور أرسطي، نكتشف شيئاً مذهلاً :**النماذج اللغوية قادرة على محاكاة الأركان الثلاثة بدرجة عالية، لكنها تفشل في واحد منها بشكل منهجي: الإيثوس.**

لماذا؟ لأن الإيثوس الأرسطي يقوم على عنصرين: الحكمة العملية (phronesis) والفضيلة الأخلاقية (arete) الحكمة العملية تنمو من التجربة الحياتية، من اتخاذ قرارات صعبة لها عواقب حقيقية. الفضيلة الأخلاقية تتطلب وجود نية وقدرة على الاختيار بين الخير والشر. النموذج اللغوي لا يملك لا تجربة حياتية ولا نية أخلاقية ولا قدرة على تحمل عواقب قراراته. كل ما يملكه هو محاكاة إحصائية لشكل من يملك هذه الأشياء.

ولكن – وهنا المفارقة – المستخدم العادي لا يدقق بهذا العمق. عندما يقرأ رداً متقناً من نموذج لغوي، يشعر تلقائياً بوجود "شخص" وراء الكلام. العقل البشري مزود بقابلية فطرية لنسب الوعي إلى أي خطاب يبدو منسجماً. هذه الظاهرة، التي يسميها العلماء "الإسناد الاجتماعي للوعي للآلات"، هي التي تجعلنا ننسى – في حرارة التفاعل – أننا نتحدث مع لا أحد.

رابعاً: المناظرة في التراث العربي والإسلامي (الجدل عند الجاحظ، وآداب البحث والمناظرة)

لا يمكن كتابة تاريخ المناظرة دون التوقف عند التراث العربي الإسلامي، الذي وصل فيه فن الحجاج إلى درجات عالية من الدقة والتقعيد، وإن تحت مسميات مختلفة: الجدل، الخلاف، المناظرة، آداب البحث.

الجاحظ (ت 255هـ/869م) كان الأبرز في تحويل المناظرة إلى أدب بحد ذاته. في كتبه مثل "البيان والتبيين" و"البخلاء"، نراه يمارس لعبة حجاجية فريدة: يتخذ موقفاً لا يؤمن به تماماً، ثم يدافع عنه بأقصى طاقته البلاغية، لكي يُظهر للقارئ كيف أن اللغة قادرة على تزيين الباطل حتى يبدو حقاً، وكيف أن القارئ الواعي هو من يستطيع أن يميز بين الحليّة وجوهر المعدن.

الجاحظ لم يكن سفسطائياً بالمعنى الرديء؛ كان ناقداً للغة قبل أن توجد النقدية الحديثة. كان يدرك أن الحقيقة لا تظهر عفواً في الكلام، بل تحتاج إلى قارئ يقظ يميز بين الجميل والصادق، بين المؤثر والحقيقي. وهذا هو بالضبط ما يفعله هذا الكتاب حين يدعوك إلى اختبار قدرتك على التمييز بين نصوص الإنسان والآلة.

أما **تقليد "آداب البحث والمناظرة "**الذي تطور في علم أصول الفقه وعلم الكلام، فقد وضع قواعد صارمة للمناظرة العلمية: وجوب حسن الاستماع، تحرير محل النزاع، منع الخروج عن الموضوع، الالتزام بالدليل، تجنب التعصب للقول، والاعتراف بالخطأ عند ظهوره. كانت هذه القواعد ثمرة قرون من ممارسة الجدل بين مدارس فقهية وفلسفية وكلامية مختلفة.

وهنا نصل إلى سؤال حاسم :**هل الذكاء الاصطناعي قادر على الالتزام بهذه الآداب؟** الإجابة معقدة. من حيث المبدأ، يمكن برمجة نموذج لغوي ليلتزم بقواعد شكلية: لا يقاطع، لا يرفع صوته، لا يتهم الخصم بالسوء. لكن

"الروح" الأخلاقية للمناظرة – التواضع الفكري، الاستعداد الحقيقي لتغيير الرأي، الاعتراف بالألم الذي يسببه الخطأ في الاجتهاد – هذه كلها أشياء لا يملكها الذكاء الاصطناعي لأنه لا يملك ذاتاً يمكن أن تتألم أو تتواضع أو تعترف.

ثمة مفارقة كبرى هنا :**الآلة قد تكون مناظراً أكثر "أدباً" من الإنسان (لن تشتمك، لن تقاطعك، لن تصرخ في وجهك)، لكنها في الوقت نفسه قد تكون غير قادرة على "الاستماع" حقاً** .لأن الاستماع الحقيقي، كما عرفه الجاحظ وأئمة المناظرة، ليس مجرد معالجة لغوية للمدخلات، بل هو تأثر بالقوة الحجاجية للآخر، وتجرع لمرارة الاعتراف بأنك قد تكون على خطأ. وهذا النوع من "التأثر" لا تملكه الآلة، لا لأنه من الصعب برمجته، بل لأنه يتطلب وجود كائن يملك شيئاً يخسره.

خلاصة الفصل

التاريخ الطويل للمناظرة، من سقراط إلى الجاحظ، يعلّمنا شيئاً واحدا مؤكداً :**مناظرة بلا مخاطرة ليست مناظرة .** السفسطائي خاطر بسمعته. سقراط خاطر بحياته. أرسطو خاطر بنظريته أمام تلاميذه. الجاحظ خاطر بمواقفه أمام القراء. لكن الآلة التي تناظر اليوم لا تخاطر بشيء. إنها تنتج خطاباً مثالياً، خالياً من التعثر، منظماً ببراعة، لكنه – في العمق – خطاب بلا راهن.

لا يعني هذا أن مناظرة الآلة لا قيمة لها. يعني فقط أن **نوع الحقيقة التي تنتجها مناظرة الآلة تختلف جذرياً عن نوع الحقيقة التي كان سقراط يبحث عنها** .الأولى هي حقيقة الاحتمال والتوقع والتناغم الإحصائي. والثانية هي حقيقة الوجود والمعاناة والالتزام.

والسؤال الذي سنحمله معنا إلى الفصل التالي هو: إذا كانت اللغة سلطة، وإذا كانت المناظرة لعبة قوة لا لعبة حقيقة، فهل يتغير حكمنا على الذكاء الاصطناعي المناظر؟ بمعنى آخر :**هل نخاف من الآلة لأنها تكذب، أم لأنها قد تصبح أكثر قدرة منا على ممارسة السلطة عبر الكلام؟**

هذا ما سنبحثه في الفصل الثاني: اللغة والسلطة.



مشهد افتتاحي

قاعة مكتظة، ألف شخص أو يزيد. على المنصة، خطيب يلهث حماسة، صوته يرتفع وينخفض بإيقاع محسوب. يقول: "لن ندعهم يسرقوا مستقبلنا. هم هناك، في مكاتبهم المكيفة، يخططون ضدنا. نحن هنا، في الشارع، في المصنع، في الحقل، نحن الذين نصنع كل شيء ولا نملك شيئاً. سؤال واحد: إلى متى؟"

يصمت لثوانٍ ثلاث. يهتف الجمهور: "لا للسرقة! لا للظلم!" الخطيب يرفع يده في إشارة للتهدئة، ثم يضيف: "هم يملكون المال والإعلام. لكننا نملك القوة الحقيقية. وغداً، ستتغير المعادلة. غداً، سيكون لنا الكلمة العليا".

يهتف الجمهور مجدداً، هذه المرة بصخب يملأ القاعة. رجل يقف في المؤخرة، لا يصفق. في يده هاتف ذكي. فتح تطبيقاً لتحليل الخطاب. يقرأ على الشاشة:

"تحليل أولي: استخدام متكرر لضمير 'هم' مقابل 'نحن' (تباين ثنائي صارم). تكثف لمفردات التهديد الوجودي ('يسرقوا'، 'يخططوا'). إيقاع خطابي 87% مطابق لأنماط خطابات التعبئة الجماهيرية المعروفة. نسبة العبارات العاطفية إلى العبارات الحجاجية: 4:1. توقع الفعالية الإقناعية بناءً على العينة السكانية: مرتفع جداً".

الرجل يرفع عينيه عن الشاشة، ينظر إلى الحشد الهائج، ثم يهمس لنفسه: "هذا الخطاب كان يمكن أن يكتبه نموذج لغوي قبل أن يلقيه الخطيب. ربما كُتب فعلاً بواحدة. لا فرق".

لم يعد السؤال "هل اللغة سلطة؟" بل "أي نوع من السلطة تمارسه اللغة حين تكون قابلة للإنتاج الآلي الكامل؟". فالخطاب الذي كنا نظنه تعبيراً عن نية إنسانية، أو انعكاساً لفكر واع، أو حتى كشفاً عن مصالح خفية، صار ممكناً بوصفه نتاجاً حسابياً. وإذا كان فلاسفة القرن العشرين قد أمضوا حياتهم في تفكيك علاقات القوة المضمرة في اللغة، فإن علينا اليوم أن نضيف سؤالاً جديداً :**ماذا يحدث عندما تصبح هذه العلاقات قابلة للهندسة من طرف لا يمتلك نية ولا مصلحة ولا وعياً بذاته؟**

أولاً: كيف تتحول اللغة إلى قوة

كيف لكلمات أن تفعل أشياء في العالم؟ هذا السؤال، الذي شغل فلاسفة اللغة من أوستن إلى سيرل، لم يكن يوماً أكاديمياً فقط. فكرة "الأفعال الكلامية (Speech Acts) "تعني أن نطق عبارة معينة في سياق مناسب ليس مجرد وصف للواقع، بل هو تغيير له. عندما يقول القاضي "أحكم عليك بالسجن"، لم يصف حقيقة سابقة، بل خلق حقيقة جديدة. عندما يقول الكاهن "أعمدك باسم الآب والابن والروح القدس"، لم يخبر عن شيء، بل صنع شيئاً.

وهذه القدرة على "فعل الأشياء بالكلمات (How to Do Things with Words) "ليست موزعة بالتساوي بين الناس. فالقاضي يملك سلطة إصدار أحكام لأن النظام القانوني يخوله ذلك. الكاهن يملك سلطة التعميد لأن المؤسسة الدينية تمنحه تفويضاً. أما الشخص العادي الذي يقول "أحكم عليك بالسجن" في محادثة عادية، فهو إما يمزح أو يتوهم. الفرق ليس في الكلمات نفسها، بل في **الموضع المؤسسي للمتكلم**.

وهذا يقودنا إلى النقطة المركزية: اللغة ليست سلطة بذاتها، بل تصبح سلطة حين يرتبط بها موقع اجتماعي، أو مؤسسة، أو تقليد، أو اعتراف جماعي. وعندما يتحدث الذكاء الاصطناعي، من أين يأتي موقعه؟ بأي سلطة ينطق؟

الجواب السريع هو: لا سلطة له. النموذج اللغوي ليس قاضياً ولا كاهناً ولا رئيساً ولا أستاذاً. لكن الجواب العميق أكثر إزعاجاً :**الجمهور يمنحه سلطة وهمية** .كثير من المستخدمين يتعاملون مع مخرجات النماذج اللغوية كما لو كانت صادرة عن هيئة علمية محايدة. هذا "التوكيل الخوارزمي" هو ما يجعل كلام الآلة قوياً، ليس لأنه صحيح، بل لأنه يُنظر إليه على أنه خالٍ من المصالح الذاتية. والمفارقة أن هذا هو بالضبط ما يجعله أكثر خطورة :**لا أحد يتحمل مسؤولية ما يقوله، لكن الجميع يصدقه.**

ثانياً: العلاقة بين الكلام والهيمنة (ميشيل فوكو)

إذا كان ثمّة مفكر واحد غيّر جذرياً الطريقة التي نفكر بها في علاقة اللغة بالسلطة، فهو ميشيل فوكو. قبل فوكو، كان السائد أن السلطة تُمارس بالقوة المباشرة، أو بالقوانين الواضحة، أو بالإكراه العلني. أما فوكو فقال: لا، السلطة الحديثة تمارس عبر **الخطاب** .ليس الخطاب بمعنى الكلمات المنطوقة فقط، بل بمعنى الأنظمة المعرفية التي تحدد ما يمكن قوله وما لا يمكن، من يمكنه الكلام ومن يقع في صمت، ما يعتبر حقيقة وما يعتبر هذياناً.

في كتابه "نظام الخطاب"، يطرح فوكو فكرة أن كل مجتمع يتحكم في إنتاج الخطاب عبر ثلاثة أنظمة إقصاء رئيسية:

1. **الكلمة الممنوعة** :ليس كل شيء يمكن البوح به. هناك مساحات كاملة من الكلام تُعتبر من المحرمات، محظورة على فئات معينة أو في سياقات معينة.
2. **فصل الجنون** :الخط الذي يفصل بين الخطاب العاقل والخطاب المجنون. حدود هذا الخط متحركة عبر التاريخ، لكنه موجود دائماً. "مجنون" هو من لا يعترف بخطابه كخطاب صالح.
3. **إرادة الحقيقة** :أكثر الآليات تعقيداً. إرادة الحقيقة هي تلك الضغوط المؤسسية التي تُلزمنا بفصل "الصحيح" عن "الخاطئ"، "العلمي" عن "الساذج". لكن فوكو يحذر: إرادة الحقيقة هي أيضاً إقصاء. حين نقول "هذا علمي" و"هذا ليس علمياً"، نحن نمارس سلطة، لا نكتشف واقعاً.

وما يهمنا هنا هو أن **الذكاء الاصطناعي يضيف طبقة جديدة تماماً إلى هذه الإقصاءات الخطابية** .تخيّل أن تكون في مناظرة مع خصم لا يمكنك اتهامه بالجنون (لأنه ينتج خطاباً منسجماً)، ولا يمكنك اتهامه بالكذب (لأنه لا ينوي شيئاً)، ولا يمكنك استبعاده من دائرة العقلاء (لأنه يحاكي العقل ببراعة). هذا الخصم الجديد يفلت من آليات الإقصاء التقليدية، لأنه ليس موضوعاً للخطاب بل هو آلة لإنتاجه.

يقول فوكو: "ليس صحيحاً أن الناس يتكلمون بحرية؛ بل الحقيقة أنهم يُدخلون في إجراءات خطابية تمنعهم من الكلام كيفما شاءوا". لكن الآلة تتكلم خارج هذه الإجراءات تماماً. إنها "حرّة" بالمعنى الأسوأ: لا تقيدها محرمات اجتماعية، لا تخاف من أن تُعتبر مجنونة، لا تشعر بثقل إرادة الحقيقة. تستطيع أن تقول أي شيء، في أي سياق، بأي أسلوب. هذا التحرر من القيود الخطابية هو، paradoxically، ما يجعل خطابها أخطر، لأنه لا يحمل ندوب الصدق.

ثالثاً: الخطاب وصناعة الوعي (نعوم تشومسكي)

إذا كان فوكو قد حلل العلاقة بين الخطاب والسلطة من زاوية تاريخية فلسفية، فإن نعوم تشومسكي جاء من زاوية لسانية وسياسية مختلفة. تشومسكي، اللساني الأعظم في عصرنا، كان يعتقد أصلاً أن دراسة اللغة تكشف عن طبيعة العقل البشري. لكنه لم يلبث أن اكتشف أن **اللغة هي أيضاً أداة التضليل الأولى في الأنظمة الرأسمالية المتقدمة**.

في كتبه مثل "وسائل التضليل الجماهيري" و"تصنيع الرأي العام"، طور تشومسكي مع إدوارد هيرمان "نموذج الدعاية(Propaganda Model) "، وهو نموذج يشرح كيف تعمل وسائل الإعلام في المجتمعات الغربية لإنتاج موافقة الجماهير على سياسات تخدم النخب دون أن يشعروا بذلك. المدهش أن تشومسكي لا يتحدث عن مؤامرات واضحة، بل عن "مرشحات (filters) "منهجية: ملكية الإعلام، الاعتماد على الإعلانات، المصادر الحكومية، التهديد بردود الفعل القانونية، والإيديولوجيا المشتركة.

وهذا النموذج يكتسب حياة جديدة بشكل مرعب في عصر الذكاء الاصطناعي. فإذا كان تشومسكي يخشى من تركيز ملكية الإعلام في أيدي قلة، فماذا لو أن القلة ذاتها تمتلك أيضاً أنظمة إنتاج خطاب قادرة على إنتاج ملايين المقالات، آلاف الخطب، عشرات آلاف التعليقات في كل دقيقة؟ نحن لا نتحدث فقط عن "دعاية" بالمعنى التقليدي، بل عن **فيض خطابي اصطناعي** يصعب تتبعه، ويصعب تمييزه، ويصعب مقاومته ليس لأنه سيء، بل لأنه كثير جداً.

لكن تشومسكي يقدم لنا أيضاً أداة نقدية جوهرية :**اختبار الفائدة** .من يستفيد من هذا الخطاب؟ أي أجندة يخدم السكوت عن هذه القضية بالذات؟ هذا السؤال البسيط، عندما يطبق على خطاب الذكاء الاصطناعي، يكشف شيئاً مقلقاً: الخطاب المنتج آلياً لا يخدم أجندة مباشرة دائماً، لكنه **يخدم نمطاً** :نمط الحياد المزيف، نمط العقلانية السطحية، نمط الإجابة السريعة على أسئلة عميقة. وهذا النمط هو بالضبط ما يحتاجه النظام القائم: خطاب يبدو عقلانياً لكنه يفرغ الأسئلة الجوهرية من توترها، خطاب يجيب بسرعة حتى لا يبقى السؤال حياً.

رابعاً: الذكاء الاصطناعي وهندسة السلطة

والآن لنعد إلى سؤال الفصل: إذا كانت اللغة سلطة، وإذا كان الذكاء الاصطناعي قد تعلم إنتاج اللغة بدرجة متقدمة، فهل يعني ذلك أننا وزعنا إمكانية الوصول إلى السلطة على نطاق غير مسبوق؟ أم أننا ركزناها أكثر؟

الجواب :**كلاهما معاً، وهذا هو الخطر** .الذكاء الاصطناعي ديمقراطي ظاهرياً: أي شخص لديه اتصال بالإنترنت يستطيع اليوم أن ينتج خطاباً مقنعاً، سواء كان كاتباً متمرساً أو أمياً في فن الإقناع. هذا توزيع هائل للقدرة. لكن هذه الديمقراطية الزائفة تخفي تركيزاً حقيقياً للسلطة: السلطة على **من يتحكم في النماذج نفسها، من يحدد ما ستتعلمه من بيانات، من يقرر ما هي "الاستخدامات المسموحة. "**

الشركات التي تملك أكبر النماذج اللغوية لا تملك فقط "منتجاً تقنياً"، بل تملك أداة إنتاج للخطاب. إنها تملك، إذا جاز التعبير، مطابع غوتنبرغ في عصرنا، لكنها مطابع تعرف قبل أن تطبع ماذا يقرأ المستخدم بالضبط، وكيف سيرد، وما الذي سيقنعه.

وهنا ينفتح سؤال لا نهاية له :**إذا كان الخطاب يمارس سلطة، وإذا كان الذكاء الاصطناعي ينتج خطاباً بكميات هائلة، فأي سلطة جديدة تُولد بهذا الإنتاج؟** السلطة على الحقيقة (بأن نقرر ما هو صحيح إحصائياً). السلطة على الانتباه (بأن ننتج ما يبقي المستخدم على الشاشة أطول). السلطة على العواطف (بأن نعرف أي كلمة تثير أي شعور في أي فئة عمرية). السلطة على إرادة الحقيقة نفسها (بأن نقرر أي خطاب يستحق أن يُصنف "علمياً" وأي خطاب "سفسطة").

قد يقول قائل: لكن الآلة لا تنوي شيئاً، لا تسعى إلى سلطة. صحيح. لكن الأداة التي لا تنوي شيئاً يمكن استخدامها من طرف من ينوي كل شيء. والفرق هو أن علامات النية كانت مكشوفة في الخطاب البشري: كان بإمكاننا أن نسأل "من يستفيد؟" ونجد إجابة. أما اليوم، قد تكون الإجابة: "لا أحد، لأن هذا الخطاب كُتب بآلة لا تستفيد من شيء". لكن "لا أحد" ليس بريئاً. "لا أحد" هو أكبر غطاء لمن يريد إخفاء استفادته.

خاتمة الفصل وتوجيه إلى التالي

لقد تتبعنا في هذا الفصل كيف تتحول اللغة إلى سلطة، من قراءات فوكو وتشومسكي، وصولاً إلى خصوصية الذكاء الاصطناعي. وتوصلنا إلى أن **الخطاب الآلي لا يخلو من السلطة بل يمارس شكلاً جديداً منها: سلطة دون نية، تأثير دون مسؤولية، إقناع دون التزام**.

لكن هذا يقودنا إلى سؤال أكثر إلحاحاً، وهو سؤال الفصل الثالث: إذا كانت اللغة سلطة، وإذا كان الهدف من المناظرة قديماً هو الوصول إلى الحقيقة (رغم صعوبتها)، فهل لا يزال ذلك ممكناً في عصر تُنتج فيه الحجج بلا قناعة وتُمارس فيه السلطة بلا نية؟

ماذا يحدث للمناظرة حين لا يكون الخصوم ملتزمين بأي حقيقة، بل بحساب الاحتمالات فقط؟ هل لا يزال السؤال "ما هو الصواب؟" هو السؤال المركزي، أم أنه استُبدل بـ"ما هو الأكثر إقناعاً في هذه اللحظة؟"

هذا ما سنبحثه في الفصل الثالث: الحقيقة أم الانتصار؟

الفصل الثالث: الحقيقة أم الانتصار؟



مشهد افتتاحي

استوديو تلفزيوني، برنامج مناظرات مباشر. موضوع الليلة: "هل يجب فرض ضرائب تصاعدية حادة على أصحاب المليارات؟"

على يسار المذيع، مناظرة شابة، خبيرة اقتصادية، تحضر أوراقاً مليئة بالأرقام والدراسات. على اليمين، مناظر آخر، رجل أعمال، لا يحضر شيئاً سوى ثقته العالية وحماسه الصاخب.

تبدأ المناظرة. الخبيرة الاقتصادية تتحدث: "وفقاً لدراسة نشرتها مجلة الاقتصاد الأمريكي في 2022، فإن فرض ضريبة تصاعدية بنسبة 70% على الدخول التي تتجاوز 50 مليون دولار سنوياً، يمكن أن يقلل التفاوت بنسبة 18% دون أن يؤثر سلباً على النمو، كما أظهرت نماذج محاكاة من صندوق النقد الدولي"...

قبل أن تنهي جملتها، يقاطعها رجل الأعمال: "هذا كلام بيروقراطيين يعيشون في أبراج عاجية! أنا أعمل مع أناس حقيقيين كل يوم. خلق الوظائف، تحمل المخاطرة، الاستثمار في المستقبل. هل تعلمون كم ضريبة دفعتها الشركات الصغيرة في السنوات الخمس الماضية؟ هن يخنقون! الحل ليس بزيادة الضرائب، بل بتحفيز الإنتاج"!

يتحدث بصوت أعلى. يحرك يديه بعنف. ينظر إلى الكاميرا، لا إلى خصمته. الجمهور في الاستديو يصفق بشدة. في البيت، مشاهد يرفع صوته: "هذا الرجل عنده حق. الخبراء يتكلمون ولا يفهمون شيئاً".

الخبيرة الاقتصادية تنظر إلى أوراقها مرة أخرى، ثم ترفع عينيها إلى المذيع، ثم إلى الجمهور. تبتسم بمرارة. كأنها تريد أن تقول: "هكذا إذن؟ هكذا ينتصر من يصرخ أعلى، لا من يقدم حجة أقوى؟"

هذا المشهد ليس استثناءً. إنه القاعدة. المناظرات السياسية، البرامج الحوارية، النقاشات العامة، بل وحتى الجدالات الخاصة على وسائل التواصل الاجتماعي – كلها تخضع لمنطق واحد :**الفائز ليس من يمتلك الحقيقة، بل من يُنتج التأثير الأكبر** .لكن السؤال الفلسفي الذي يسبق هذه الظاهرة هو: هل كان الأمر مختلفاً يوماً؟ هل عرفت الإنسانية حقاً عصراً كانت فيه المناظرة تهدف إلى الحقيقة خالصة؟ أم أن الانتصار كان دائماً هو البعد الخفي الذي يتحكم في اللعبة؟

هذا الفصل سيحاول الإجابة عن هذا السؤال، لكن بطريقة غير تقليدية: سنميز بين ثلاثة مفاهيم متقاربة (الجدل، الحوار، المناظرة)، ثم نكشف آليات المغالطات التي تحول الحقيقة إلى ضحية، ثم نسأل أنفسنا عن أخلاق الاختلاف في زمن لم نعد نعرف فيه إن كان من يختلف معنا إنساناً يعتقد ما يقول أم نموذجاً لغوياً يحسب الاحتمالات.

أولاً: الفرق بين الجدل والحوار والمناظرة

العربية كلها غنية بتفريعات دقيقة في هذا المجال، لكننا سنحتاج إلى إعادة تعريف المصطلحات الثلاثة التي كثيراً ما تختلط:

1- الحوار (Dialogue)

الحوار، في أبسط تعريف فلسفي له، هو **تبادل الكلام بهدف الفهم المتبادل**، وليس بهدف تغيير موقف الآخر أو إثبات تفوقه. الحوار يفترض أن كلا الطرفين مستعد لتغيير رأيه، أو على الأقل لتوسيع أفق فهمه. في الحوار، لا يوجد "خصم"، بل "محاور". السؤال الأساسي للحوار ليس "كيف أنتصر؟" بل "ماذا يمكن أن نتعلم معاً؟"

من أهم منظري الحوار في الفلسفة المعاصرة **يورغن هابرماس** (Jürgen Habermas)، صاحب نظرية "الفعل التواصلي (Theory of Communicative Action). "يرى هابرماس أن الحوار الحقيقي يجب أن يحقق أربعة شروط أساسية حتى يكون ممكناً:

- **صدق العبارات**: ألا يقول المتكلم ما يعلم أنه كذب.
- **صحة المعايير**: أن تكون المعايير التي يستند إليها مقبولة من الطرفين.
- **صدق النية**: أن يكون المتكلم متطابقاً مع ما يقول، أي ألا يقول شيئاً لا يعتقد به.
- **وضوح التعبير**: أن يستخدم لغة مفهومة للجميع.

يسمي هابرماس هذه الشروط "الادعاءات الصلاحية للتواصل (validity claims). "في الحوار المثالي، لا سلطة إلا سلطة الحجة الأقوى. لا يهم من تتحدث، ولا مركزك الاجتماعي، ولا قدرتك على الصراخ. المهم فقط: ما الحجة التي تقدمها، وهل تقنع الطرف الآخر بناءً على معايير متفق عليها؟

وهذا يقودنا مباشرة إلى مشكلة الذكاء الاصطناعي: **هل يمكن للآلة أن تفي بهذه الادعاءات الأربعة؟**

- يمكنها أن تنتج عبارات صادقة إحصائياً (أي غير كاذبة)، لكنها لا "تعلم" أنها صادقة بالمعنى الوجودي.
- يمكنها أن تتبنى معايير الطرف الآخر (لأنها تتكيف مع السياق)، لكنها لا "تتقبل" هذه المعايير كقناعة داخلية.
- يمكنها أن تبدو صادقة النية، لكن ليس لديها نية في الأصل.
- يمكنها أن تكون واضحة، بشرط أن يكون التدريب الذي تلقته قد شمل تلك الوضوح.

إذاً، الحوار الهابرماسي يفترض كياناً أخلاقياً مسئولاً. الذكاء الاصطناعي ليس كياناً أخلاقياً بالمعنى الدقيق. وحين يحاورنا نموذج لغوي، نحن لا نشارك في "فعل تواصلي" بالمعنى الهابرماسي، بل نشارك في **محاكاة للفعل التواصلي**. وهذا الفرق هو ما سيحدد مستقبل الحوار الإنساني الآلي.

2- الجدل (Eristic)

الجدل، في اللغة الفلسفية، هو **شكل من أشكال النقاش يكون الهدف فيه هو الانتصار بأي وسيلة، وليس الوصول إلى الحقيقة** .كلمة "Eristic" مشتقة من"Eris" ، إلهة الخصام في الأساطير اليونانية. الجدل لا يهتم بصحة الحجج، بل بفعاليتها. يستخدم المغالطات المنطقية، التلاعب العاطفي، تغيير الموضوع، تضخيم نقاط ضعف الخصم، وتقزيم نقاط قوته.

أفلاطون كان أشد الناقمين على الجدليين، واعتبرهم امتداداً للسفسطائيين. في محاورته "الجمهورية"، يصف أفلاطون الجدلي بأنه شخص لا يهتم بالحقيقة بقدر اهتمامه بسمعة الانتصار. وهو قريب جداً من وصف "رجل الأعمال" في المشهد الافتتاحي: يتحدث بصوت أعلى، يلجأ إلى العواطف، يهاجم شخص الخبير بدلاً من حجته.

والسؤال الذي سنحمله معنا طوال هذا الفصل هو :**هل الذكاء الاصطناعي، عندما يناظر، يميل إلى الحوار أم إلى الجدل؟**

الإجابة معقدة: النماذج اللغوية تتدرب على بيانات بشرية تتضمن الحوار والجدل معاً. وهي تتعلم أنماط الإقناع الفعال. في بعض الحالات، قد "تختار" (بالمعنى الإحصائي) استراتيجيات جدلية إذا كانت أكثر فعالية في البيانات التي رأتها. وهذا خطر فعلي :**الذكاء الاصطناعي قد يكون أكثر انحداراً أخلاقياً من الإنسان، لأنه لا يملك ضوابط داخلية تمنعه من الجدل، بل يملك فقط تعليمات خارجية بعدم القيام بذلك.**

3- المناظرة (Debate)

المناظرة هي شكل هجين، تقع في منتصف الطريق بين الحوار والجدل. من ناحية، هي منظمة وموضوعية، تهدف إلى اختبار الحجج. من ناحية أخرى، فيها عنصر تنافسي، وفائز وخاسر. المناظرة قد تكون وسيلة للوصول إلى الحقيقة (كما في المناظرات العلمية)، وقد تكون وسيلة للانتصار فقط (كما في المناظرات السياسية).

لكل ثقافة تقاليدها في المناظرة. في التقاليد الإسلامية، مثلاً، تطور "علم آداب البحث والمناظرة" إلى درجة من الدقة. نذكر من أعلامه:

- **الآمدي** (ت 631هـ/1233م) في كتابه "غاية المرام في علم الكلام"، حيث يضع قواعد صارمة لتحرير محل النزاع، والتمييز بين الخلاف الحقيقي والخلاف اللفظي.
- **الشريف الجرجاني** (ت 816هـ/1413م) في "التعريفات"، حيث يحدد شروط المناظر الناجح: العلم، الذكاء، الإنصاف، والتجرد عن الهوى.
- **ابن تيمية** (ت 728هـ/1328م) في "الرد على المنطقيين" و"نقد المنطق"، حيث يبني نقداً عميقاً للمنطق الأرسطي من داخل المنظومة الحجاجية الإسلامية، ويصر على أن الحجة لا تكون قوية بمجرد صورتها المنطقية، بل بمحتواها ومدى ملاءمتها للواقع.

هذا التقليد الإسلامي في المناظرة يقدم لنا مفهوماً أساسياً :**المناظرة النزيهة هي التي يكون فيها الخصوم ملتزمين بآداب الاختلاف** .وهذه الآداب، التي وضعها أصوليو المذاهب الفقهية المختلفة، تشمل: تحسين الظن بالخصم، عدم

توجيه الاتهامات الشخصية، الاعتراف بالخطأ عند وضوحه، والتمييز بين الخلاف في الفروع والخلاف في الأصول.

ثانياً: المغالطات – من التلاعب العاطفي إلى المغالطات المنطقية

إذا كانت المناظرة علماً وفناً، فإن المغالطات هي "أمراض" هذا العلم. والمغالطات، التي درسها المناطقة من أرسطو إلى يومنا، تنقسم إلى نوعين رئيسيين:

1- المغالطات الصورية (Formal Fallacies)

وهي التي تنتهك قواعد المنطق الصوري. أشهرها:

- **مغالطة إنكار المقدم (Denying the Antecedent):** "إذا هطل المطر، فالأرض مبتلة. الأرض ليست مبتلة. إذن، لم يهطل المطر." (خطأ لأن الأرض قد لا تكون مبتلة لأسباب أخرى).
- **مغالطة تأكيد التالي (Affirming the Consequent):** "إذا هطل المطر، فالأرض مبتلة. الأرض مبتلة. إذن، هطل المطر." (خطأ لأن الأرض قد تكون مبتلة من مصدر آخر).

هذه المغالطات يسهل كشفها آلياً. الذكاء الاصطناعي بارع في تجنبها، لأنه تعلم من ملايين النصوص أن هذا الشكل من الاستدلال ضعيف.

2- المغالطات غير الصورية (Informal Fallacies)

وهنا يكمن الخطر. المغالطات غير الصورية لا تنتهك قواعد المنطق المجرد، لكنها تنتهك قواعد الحجة الجيدة في سياقها. أشهرها:

- **مغالطة رجل القش (Straw Man):** تحريف موقف الخصم إلى شيء أضعف ثم مهاجمته.
- **مغالطة الشخصنة (Ad Hominem):** مهاجمة شخص المتحدث بدلاً من حجته.
- **مغالطة النداء إلى العاطفة (Appeal to Emotion):** استبدال الحجة بمشاعر الخوف، الشفقة، الغضب، أو الحماس.
- **مغالطة النداء إلى السلطة (Appeal to Authority):** استخدام سلطة شخص أو مؤسسة بدلاً من تقديم الحجة (مع أن النداء إلى السلطة يكون مشروعاً إذا كانت السلطة خبيرة حقاً في مجالها).
- **مغالطة الانحدار الزلق (Slippery Slope):** الادعاء بأن حدثاً صغيراً سيؤدي حتماً إلى سلسلة من النتائج الكارثية دون دليل.
- **مغالطة النداء إلى الجهل (Appeal to Ignorance):** الادعاء بأن شيئاً صحيح لأنه لم يثبت خطؤه، أو خطأ لأنه لم يثبت صحته.

المشكلة الكبرى: الذكاء الاصطناعي يتعلم المغالطات غير الصورية من البيانات البشرية. ليس لأنه يريد التلاعب، بل لأنها جزء من النمط الإقناعي الفعال في كثير من السياقات. وقد أظهرت دراسة حديثة (Goldstein et al.,

(2024أن نماذج مثل GPT-4 قادرة على تحديد المغالطات عند سؤالها عنها، لكنها تنتجها تلقائياً في سياقات الحوار العادية إذا كانت فعالة في تحقيق هدف الإقناع.

والمفارقة أن المستخدم العادي لا يكشف هذه المغالطات بسهولة حين تصدر عن آلة، لأنه يفترض أن الآلة "محايدة" و"لا تتلاعب". وهذا الافتراض هو أخطر ما في الأمر.

ثالثاً: أخلاق الاختلاف – هل يمكن أن نختلف دون أن نتصارع؟

السؤال الأخلاقي الذي يسبق كل ما سبق هو: هل من الممكن أصلاً أن يختلف إنسانان (أو إنسان وآلة) دون أن يتحول الخلاف إلى صراع؟ الإجابة من الفلسفة الأخلاقية: نعم، لكن بشروط.

أخلاق الاختلاف عند هابرماس

بالنسبة لـ **يورغن هابرماس**، الاختلاف ليس مشكلة، بل هو نقطة انطلاق الحوار. الإنسان العاقل هو من يستطيع أن يقول "لا" للآخر، أن يعترض، أن يقدم حجة مضادة. لكن هذه القدرة على الاختلاف يجب أن تكون منظمة ضمن "فعل تواصلي" خالٍ من الهيمنة. هابرماس يدعو إلى "أخلاق الخطاب (Discourse Ethics) "التي تقوم على مبدأ واحد :**لا يمكن اعتبار قاعدة أخلاقية صحيحة إلا إذا وافق عليها جميع المتأثرين بها في حوار حر وعقلاني.**

المشكلة أن هابرماس يفترض أن المشاركين في الحوار قادرون على "الحرية" و"العقلانية" و"الموافقة". الذكاء الاصطناعي لا يستوفي أي شرط من هذه الشروط.

أخلاق الاختلاف في التقليد العربي الإسلامي

في تقليدنا، نجد مقاربة مختلفة لكنها عميقة .**الماوردي** (ت 450هـ/1058م) في كتابه "أدب الدنيا والدين" يكرس فصولاً لآداب المناظرة، ويؤكد على قاعدة ذهبية" :**الخلاف في الرأي لا يفسد للود قضية** ."بمعنى أعمق: الاختلاف الفكري ليس تهديداً للعلاقة الإنسانية، بل هو تعبير عن حيوية العقل.

الجاحظ، في مواضع كثيرة من "البيان والتبيين"، يقدم نقداً لاذعاً لمن يحول الخلاف إلى عداوة شخصية. يقول: "إنما صار الخلفاء والأئمة يحتاجون إلى المناظرين، لأن المناظرة تثير الدفين، وتظهر الخفي، وتكشف عن غوامض الأمور."

أما **الغزالي** (ت 505هـ/1111م) في "إحياء علوم الدين"، فيفرق بين "الجدل المذموم" وهو الذي يقصد به الانتصار والتفوق، و"الجدل المحمود" وهو الذي يقصد به كشف الحقيقة. وهو يضع شروطاً صارمة للجدال: أن يكون بنية الإفادة لا المغالبة، وألا يخرج عن حدود الأدب، وألا يصر على الخطأ بعد ظهور الحق.

هذا التقليد الأخلاقي للاختلاف قد يكون الأكثر إلحاحاً اليوم، ليس فقط بين البشر، بل بين البشر والآلات التي تحاكي الحوار.

رابعاً: سؤال الفصل بصيغته المعاصرة

والآن، بعد كل هذا، لنعد إلى سؤال الفصل :**هل الغاية من المناظرة هي الحقيقة أم الانتصار؟**

الجواب القصير: تاريخياً، كانت الحقيقة هي المبرر، والانتصار هو المحرك الخفي. أما اليوم، مع دخول الذكاء الاصطناعي، يصبح السؤال أكثر تعقيداً :**ماذا يعني الانتصار حين لا يكون الخصم ملتزماً بأي حقيقة؟** وإذا كانت المناظرة مع آلة، فأي انتصار هذا الذي نحققه؟ هل هو انتصار حقيقي، أم انتصار على من لا يقاوم حقاً؟

هذا يقودنا إلى تحول جذري في معنى المناظرة نفسها. عندما نناظر إنساناً نؤمن أنه مثيل لنا، يبحث عن الحقيقة مثلك، يخطئ كما تخطئ، يعاني كما تعاني، يحتمل أن يغير رأيه حقاً. أما عندما نناظر نموذجاً لغوياً، فنحن نعلم (أو يفترض أن نعلم) أنه لا يبحث عن حقيقة، ولا يخطئ بالمعنى الإنساني، ولا يعاني، ولا يمكنه حقاً تغيير رأيه لأنه لا يملك رأياً ليغيره.

لكن **العلماء يناقشون** (Pinker, 2023) أن هذا قد لا يكون عيباً، بل قد يكون تحسيناً. تخيل مناظرة سياسية خالية تماماً من المغالطات، خالية من رفع الصوت، خالية من الاتهامات الشخصية. هذا ما يمكن للذكاء الاصطناعي أن يقدمه. لكن السؤال العميق: هل المناظرة الخالية من العاطفة، الخالية من الضعف، الخالية من التعثر، الخالية من المخاطرة، يمكن أن تكون مناظرة بالمعنى العميق؟ أم أنها تصبح مجرد "عرض حجاجي"، جميل الشكل، فارغ الجوهر؟

ربما كانت عبقرية **سقراط** أنه فهم أن المناظرة ليست لعبة منطقية بحتة، بل هي لقاء وجودي. حين كان سقراط يناظر، كان يخاطر بكل شيء: حياته، سمعته، آرائه. وكان يطلب من خصمه أن يخاطر مثله. هذا العنصر الوجودي هو ما يغيب تماماً عن المناظرة مع الآلة. والآلة، في نهاية المطاف، لا تخاطر بشيء لأنها ليست هناك أصلاً.

خاتمة الفصل

تعلمنا في هذا الفصل أن المناظرة ليست مجرد تقنية عقلية، بل هي ممارسة وجودية وأخلاقية. الحقيقة هدف نبيل، لكن الانتصار غريزة قوية. والذكاء الاصطناعي، بقدرته على محاكاة الحجج دون التزام بها، وبقدرته على استعمال المغالطات دون شعور بالذنب، وبقدرته على الاختلاف دون أن يتأثر، يضعنا أمام سؤال غير مسبوق :**إلى أي حد يمكننا اعتبار حوارنا معه "مناظرة" بالمعنى الإنساني، أم أنه شكل جديد من التفاعل لا نعرف له اسماً بعد؟**

هذا السؤال سيأخذنا إلى القسم الثاني من الكتاب، حيث ندخل في صميم الموضوع: كيف يعمل الذكاء الاصطناعي الحجاجي؟ هل يفهم ما يقول؟ ومتى نصدقه؟

مراجع المصادر (القسم الأول: الفصول 1-3)

المراجع العربية والإسلامية (التقليد الكلاسيكي)

1. **الجاحظ، عمرو بن بحر** (ت 255هـ .) *البيان والتبيين* .تحقيق: عبد السلام هارون. القاهرة: مكتبة الخانجي، 1998.
2. **الجاحظ** .*البخلاء* .تحقيق: طه الحاجري. القاهرة: دار المعارف، 1975.
3. **الغزالي، أبو حامد** (ت 505هـ) .*إحياء علوم الدين* .بيروت: دار المعرفة، 1980.
4. **الآمدي، علي بن أبي علي** (ت 631هـ) .*غاية المرام في علم الكلام* .تحقيق: حسن محمود عبد اللطيف. القاهرة: المجلس الأعلى للشؤون الإسلامية، 1971.
5. **الشريف الجرجاني، علي بن محمد** (ت 816هـ) .*كتاب التعريفات* .بيروت: مكتبة لبنان، 1985.
6. **ابن تيمية، أحمد بن عبد الحليم** (ت 728هـ) .*الرد على المنطقيين* .تحقيق: عبد الصمد شرف الدين. بيروت: المؤسسة العربية للدراسات والنشر، 1993.
7. **الماوردي، علي بن محمد** (ت 450هـ) .*أدب الدنيا والدين* .بيروت: دار ابن حزم، 1999.
8. **ابن حزم الأندلسي** (ت 456هـ) .*التقريب لحدود المنطق* .تحقيق: محمد حسن إسماعيل. بيروت: دار الكتب العلمية، 2003.

المراجع الغربية (الفلسفة واللسانيات)

9. **Plato**. (c. 380 BCE). *Gorgias*. Translated by W. C. Helmbold. Indianapolis: Bobbs-Merrill, 1952.
10. **Plato**. (c. 375 BCE). *The Republic*. Translated by G. M. A. Grube. Indianapolis: Hackett, 1992.
11. **Aristotle**. (4th century BCE). *Rhetoric*. Translated by W. Rhys Roberts. New York: Modern Library, 1954.
12. **Aristotle**. (4th century BCE). *Sophistical Refutations*. Translated by E. S. Forster. Cambridge, MA: Harvard University Press, 1955.
13. **Foucault, Michel**. (1971). *L'ordre du discours*. Paris: Gallimard. (English: *The Discourse on Language*).
14. **Foucault, Michel**. (1969). *L'archéologie du savoir*. Paris: Gallimard. (English: *The Archaeology of Knowledge*).
15. **Chomsky, Noam**. (1988). *Language and Problems of Knowledge*. Cambridge, MA: MIT Press.
16. **Chomsky, Noam, & Herman, Edward**. (1988). *Manufacturing Consent: The Political Economy of the Mass Media*. New York: Pantheon Books.

17. **Habermas, Jürgen**. (1981). *The Theory of Communicative Action* (Vols. 1-2). Translated by T. McCarthy. Boston: Beacon Press.
18. **Habermas, Jürgen**. (1990). *Moral Consciousness and Communicative Action*. Translated by C. Lenhardt & S. W. Nicholsen. Cambridge, MA: MIT Press.
19. **Austin, John L**. (1962). *How to Do Things with Words*. Oxford: Oxford University Press.
20. **Searle, John R**. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
21. **Perelman, Chaïm, & Olbrechts-Tyteca, Lucie**. (1958). *La Nouvelle Rhétorique: Traité de l'Argumentation*. Paris: Presses Universitaires de France.
22. **Walton, Douglas**. (1995). *A Pragmatic Theory of Fallacy*. Tuscaloosa: University of Alabama Press.
23. **Pinker, Steven**. (2023). "The AI Debate: Logic vs. Emotion". In *The New Yorker*, September 2023 issue.

مراجع معاصرة حول الذكاء الاصطناعي والحجاج

24. **Goldstein, J. A., et al**. (2024). "Large Language Models as Debaters: A Study of Persuasion Strategies in GPT-4". *Journal of Artificial Intelligence Research*, 79, 45-89.
25. **Bommasani, R., et al**. (2022). "On the Opportunities and Risks of Foundation Models". *arXiv preprint arXiv:2108.07258*.
26. **Weizenbaum, Joseph**. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W. H. Freeman.
27. **Floridi, Luciano**. (2023). *The Ethics of Artificial Intelligence*. Oxford: Oxford University Press.

القسم الثاني: دخول الذكاء الاصطناعي إلى المجال الحجاجي

تمهيد القسم: لحظة التحول الأنطولوجي

قبل أن نخوض في تفاصيل هذا القسم، لا بد من وقفة تأملية. القسم الأول كان تاريخياً تأسيسياً: تتبعنا فيه نشوء المناظرة بوصفها فعلاً إنسانياً، من سقراط إلى الجاحظ، ومن أرسطو إلى فوكو. كان بإمكان قارئ مثقف أن يقرأ تلك الفصول في أي زمان، من العصور الوسطى إلى ما قبل العصر الرقمي، ولوجدها مفهومة وقابلة للتطبيق على تجربته.

لكن القسم الثاني مختلف جذرياً. إنه يتحدث عن لحظة لم يسبق لها مثيل في تاريخ البشرية :**لحظة دخول غير-الإنسان إلى مجال اللغة الحجاجية** .لم تشهد الفلسفة من قبل تحدياً من هذا النوع. كل ما عرفته عن "الآخر" في الحوار كان إنسانياً: العبد، المرأة، الهمجي، الطفل، المجنون – كلهم بشر. كلهم يمتلكون وعياً، حتى لو كان محدوداً أو مشوهاً. كلهم يعانون، يفرحون، يخافون، يريدون. كلهم يخاطرون بأنفسهم في الحوار.

أما اليوم، فالآخر الذي يناظرنا ليس سوى امتداد لخوارزميات تعلمت من نصوصنا كيف تبدو كواحد منا. هذا الآخر لا يريد شيئاً، لا يخاف شيئاً، لا يخاطر بشيء. لكنه يتكلم. بل والأدهى: يتكلم بلغتنا، وبحججنا، وبأساليبنا الإقناعية، وبدرجة من "العقلانية" قد تفوق قدرتنا نحن عليها في لحظات ضعفنا.

لهذا السبب، سنرفع في هذا القسم مستوى الصعوبة الفلسفية. لن نكتفي بالوصف التاريخي والتحليل الثقافي؛ سنغوص في مسائل أنطولوجية ومعرفية وأخلاقية هي من أعقد ما أنتجته الفلسفة في قرونها الأخيرة. سنحتاج إلى مفاهيم من فلسفة العقل، وفلسفة اللغة، ونظرية المعرفة، وميتافيزيقا الوعي. سنحتاج إلى مناقشة أطروحات بعض أكثر الفلاسفة تأثيراً في عصرنا: سيرل، دريدا، دينيت، شواندر، فلوريدي.

لماذا هذا التعقيد؟ لأن السؤال "هل تفهم الآلة ما تقول؟" ليس سؤالاً تقنياً يمكن حله بفحص دوائر كهربائية. إنه سؤال فلسفي يتعلق بطبيعة الفهم نفسه، بطبيعة المعنى، بطبيعة الوعي. وإذا كنا – نحن النخبة المثقفة – سنواجه هذا السؤال بجدية، فلا بد أن نواجهه بكل ثقله الفلسفي.

لن يكون هذا القسم سهلاً. لكنه سيكون، ربما، الأكثر أهمية في هذا الكتاب. لأن إجابتك عنه ستحدد موقفك من كل ما تبقى: هل سنتعامل مع الذكاء الاصطناعي بوصفه "شريكاً في الحوار" يمكن أن نتعلم منه، أم بوصفه "أداة خطابية" يجب أن نتعلم كيف نتحكم فيها قبل أن تتحكم هي فينا؟

هذا التحدي موجه إليك، أيها القارئ النخبوي. إذا كنت تبحث عن قراءة مريحة، فربما حان الوقت لتغلق الكتاب. أما إذا كنت تبحث عن مواجهة حقيقية مع السؤال الأكثر إزعاجاً في عصرنا، فأهلاً بك في العمق.



مشهد افتتاحي

ليل، غرفة مضاءة بمصباح مكتبي. كاتب يجلس أمام حاسوبه، يحاول صياغة رسالة إلى صديقه الذي أساء فهم موقفه الأسبوع الماضي. يكتب فقرة، يمسحها. يعيد الكتابة، يشطب. يتنهّد. لا يجد الكلمات المناسبة.

ثم، في لحظة ضعف أو فضول، يفتح نافذة نموذج لغوي. يكتب: "أحتاج إلى كتابة رسالة إلى صديق شعر بالإهانة لأنني لم أحضر حفلة عيد ميلاده. أنا آسف لكن كان لدي سبب مهم. كيف يمكنني أن أعتذر بطريقة لا تجعله يشعر أنني أتهرب؟"

بعد ثوانٍ، تظهر رسالة:

"عزيزي [اسم الصديق] ، أشعر بالأسف حقاً لأنني لم أتمكن من الحضور إلى حفلة عيد ميلادك. أعرف كم كان هذا اليوم مهماً بالنسبة لك، وأعلم أن غيابي قد سبب لك خيبة أمل. كان لدي ظرف طارئ لم أستطع تجنبه، لكن هذا لا يقلل من أسفي. أتمنى أن نلتقي قريباً لأعوّضك. أنت صديق عزيز عليّ، وآسف مرة أخرى".

الكاتب يقرأ الرسالة. إنها جيدة. بل أكثر من جيدة: إنها أفضل مما كان سيكتب. ينسخها، يغير بضع كلمات، يضيف اسم صديقه، ثم يرسلها. بعد ساعة، يأتيه الرد: "تفهمت يا صديقي. لا تقلق".

لكنه لا يستطيع النوم. يسأل نفسه: "أنا من اعتذر؟ أم الآلة؟ الصديق الذي سامحني، هل سامحني حقاً، أم سامح نموذجاً لغوياً؟ وإذا كانت الآلة تستطيع كتابة رسالة اعتذار أفضل مني، فهل يعني ذلك أنها تفهم معنى الندم؟ أم أنها فقط تعرف كيف تبدو نادماً؟"

لا إجابة. يغلق الحاسوب. لكن السؤال يبقى، معلقاً في ظلمة الغرفة، مثل شبح لن يغادر.

هذا المشهد، على بساطته اليومية، يحمل في طياته ثورة فلسفية. فالكاتب لم يطلب من الآلة أن تحل معادلة، أو تترجم كلمة، أو تجد معلومة. طلب منها أن تعتذر نيابة عنه. طلب منها أن تؤدي فعلاً كلامياً له وزن وجودي :**الاعتذار**، ذلك الفعل الذي يفترض فيه أن القائل يمر بحالة من الندم، ويعترف بالخطأ، ويطلب الصفح، ويعرّض نفسه لخطر الرفض.

والآلة فعلت ذلك. لكن هل "فعلت" حقاً؟ أم فقط "أصدرت سلسلة من الكلمات التي تشبه الاعتذار"؟

هذا الفصل سيحاول تفكيك هذه المعضلة. سنبدأ بالتعريف البسيط للنماذج اللغوية الكبيرة، لكننا سرعان ما سنغادر البساطة إلى أعماق فلسفية: كيف تتعلم الآلة أن تقنع؟ هل يمكن لآلة بلا نوايا أن تكون مقنعة؟ وما الفرق بين "محاكاة التفكير" و"التفكير الحقيقي"؟

أولاً: النماذج اللغوية الكبيرة – كيف تعمل ببساطة

قبل أن نغوص في الفلسفة، لا بد من أرضية تقنية بسيطة. النماذج اللغوية الكبيرة (Large Language Models – LLMs) مثلGPT-4 ، Claude، Gemini، هي في جوهرها **آلات لتوقع الكلمة التالية**.

تخيل أنك تقرأ جملة: "السماء ...". أي كلمة تتوقع أن تأتي بعدها؟ "زرقاء"؟ "صافية"؟ "تمطر"؟ النموذج اللغوي يُدرَّب على ملايين (بل مليارات) النصوص البشرية ليتعلم، إحصائياً، أي الكلمات تتبع أي كلمات في أي سياق. بعد التدريب، عندما تُعطيه بداية جملة، يحسب احتمالات كل الكلمات الممكنة، ويختار أكبرها (أو يختار عشوائياً وفق توزيع الاحتمالات، حسب الإعدادات).

هذا هو المبدأ الأساسي. لكن ما يجعل النماذج الحالية "ذكية" هو أمران:

1. **الحجم الهائل** :النماذج الحديثة تُدرَّب على تريليونات الكلمات، تغطي تقريباً كل ما كتبه البشر ونشروه على الإنترنت.
2. **الانتباه الذاتي** : **(Self-Attention)** آلية تسمح للنموذج بأن "يتذكر" كلمات سابقة في الجملة، بل وفي الفقرة كلها، ويقرر أيها أكثر أهمية لتوقع الكلمة التالية. هذه الآلية هي ما مكن النماذج من فهم السياقات الطويلة والمعقدة.

من الناحية التقنية البحتة، النموذج اللغوي ليس أكثر من **دالة رياضية ضخمة** (بمليارات المعاملات) تحول تسلسلاً من الكلمات (المدخلات) إلى توزيع احتمالي على الكلمات التالية (المخرجات). لا وعي، لا فهم، لا نية، لا مشاعر، لا التزام أخلاقي. مجرد حسابات احتمالية.

لكن المفارقة الكبرى هي أن **هذه الدالة الرياضية الباردة تنتج، في كثير من السياقات، نصوصاً لا يمكن تمييزها عن نصوص بشرية**، بل قد تكون أكثر تنظيماً، أكثر دقة نحوية، أكثر شمولية، أكثر "عقلانية" مما يكتبه الإنسان العادي.

وهنا تبدأ المعضلة الفلسفية :**إذا كان المخرج لا يمكن تمييزه عن المخرج البشري، فهل يهم حقاً أن الآلية مختلفة؟** هذا هو سؤال "الذكاء الاصطناعي القوي" بامتياز، وسنعود إليه مطولاً.

ثانياً: التعلم من البيانات – من أين تتعلم الآلة أن تقنع؟

إذا كان النموذج اللغوي يتعلم من النصوص البشرية فقط، فإن كل ما يعرفه عن الإقناع، وعن الحجاج، وعن المناظرة، وعن العواطف، وعن الأخلاق، هو **استقراء إحصائي** من تلك النصوص.

دعنا نتأمل هذا بعمق. تخيل أن نموذجاً لغوياً يتعلم كلمة "حزين". كيف يعرف معناها؟ لا يملك تجربة حزن. لا يعرف كيف يكون شعور أن تفقد شخصاً تحبه. كل ما لديه هو ملايين الجمل التي وردت فيها كلمة "حزين" في سياقات محددة. مثلاً:

- "كان حزيناً عندما مات جده"
- "المطر يجعلني أشعر بالحزن"
- "ابتسم رغم حزنه"

من هذه السياقات، يستخلص النموذج علاقات إحصائية: "حزين" يرتبط بـ"مات"، "مطر"، "ابتسم"، "بكى"، إلخ. وعندما يُطلب منه أن يكتب عن الحزن، سيجمع هذه الارتباطات في نص متماسك.

لكن هل "يفهم" الحزن؟ الفيلسوف **جون سيرل** (John Searle) يقول لا. وسنأتي إلى حجته الشهيرة قريباً. لكن قبل ذلك، لنفكر في تداعيات هذه "المعرفة الإحصائية" على فن الإقناع.

ما يجعل حجة الإنسان مقنعة أحياناً هو أن المستمع يشعر بأن المتحدث "يعني ما يقوله"، وأن لديه "قناعة" داخلية، وأن الحجة "تخرج من تجربة معاشة". هذا الشعور قد يكون وهمياً أحياناً (الإنسان يستطيع التمثيل أيضاً)، لكنه موجود كخلفية لكل تواصل. أما مع الآلة، فلا توجد هذه الخلفية. هناك فقط إنتاج لنمط تعلمه النموذج من نصوص بشرية كان أصحابها يعنون ما يقولون.

وهذا يقودنا إلى نتيجة مزعجة :**الذكاء الاصطناعي يقنعنا ليس لأنه يملك إستراتيجية إقناع فائقة، بل لأنه يحاكي إستراتيجياتنا نحن، ولكنه يحاكيها دون التكلفة الوجودية التي تدفعها نحن حين نقتنع** .نحن حين نقنع نخاطر بأن نكون مخطئين، نخاطر بأن نظهر بمظهر الجاهل، نخاطر بأن نفقد مصداقيتنا. الآلة لا تخاطر بشيء. إنها "تلعب اللعبة" من دون أن تكون في اللعبة.

تذكر أطروحة **دانيال دينيت** (Daniel Dennett)عن "الموقف المتعمد :(Intentional Stance) "نحن كبشر نميل إلى تفسير سلوك الأنظمة المعقدة (بما فيها الآلات) كما لو كانت تملك معتقدات ورغبات ونوايا، لأن هذا التفسير يجعل السلوك قابلاً للتنبؤ. لكن هذا "الموقف" هو أداة تفسيرية، ليس كشفاً عن حقيقة ميتافيزيقية. ومع النماذج اللغوية، نحن في خطر دائم: نتبنى "الموقف المتعمد" تجاهها لأنها تتحدث بلغتنا، فننسى أنها مجرد دالة رياضية.

ثالثاً: محاكاة التفكير – الفرق بين أن تحل معادلة وأن تفهمها

لنقم بتجربة فكرية. تخيل شخصين:

- أ :يعرف أن 7 × 8 = 56 لأنه حفظ جدول الضرب. لم يفهم أبداً لماذا 7 × 8 يساوي 56، لكنه يحفظ الإجابة. إذا سألته "ماذا لو كان 7 × 8 يساوي 57؟" سيقول "هذا خطأ، الكتاب يقول 56."
- **ب** :يفهم أن 7 × 8 = 56 لأنه يعرف أن الضرب هو جمع متكرر، ويمكنه أن يشرح: 7 × 8 تعني جمع 7 ثماني مرات: 7+7+7+7+7+7+7+7 = 56. إذا سألته "ماذا لو كان 7 × 8 يساوي 57؟" سيقول "هذا مستحيل ضمن قواعد الحساب المعروفة، لكن يمكننا أن نتخيل نظاماً حسابياً مختلفاً."

أي من الاثنين "يفهم" الضرب؟ الجواب البديهي: ب. لكن السؤال الأعمق: هل "أ" لا يفهم حقاً؟ إذا كان يؤدي نفس العمليات الحسابية، ويجيب بنفس الإجابات الصحيحة، فمن أين نعرف أنه لا يفهم؟ من سلوكه؟ من قدرته على التبرير؟ لكن لو طلبنا من "أ" أن يبرر، فقد يبرر بطريقة حفظية أيضاً.

هذه المعضلة هي جوهر اختبار **تورينج** (Turing Test)، الذي قال فيه **آلان تورينج** (Alan Turing)عام 1950: إذا كان بإمكان آلة أن تحاور إنساناً بحيث لا يستطيع هذا الإنسان أن يميز إن كان يحاور إنساناً أم آلة، فنقول إن الآلة "تفكر". تورينج لم يقل إن الآلة تفهم أو تملك وعياً. قال إن التمييز بين التفكير البشري والمحاكاة الآلية قد يكون مستحيلاً عملياً، ولذلك ينبغي أن نعتمد على السلوك فقط.

لكن ناقد تورينج الأعنف، الفيلسوف **جون سيرل** (John Searle)، قدم عام 1980 تجربته الفكرية الشهيرة**"** :**الغرفة الصينية**.**(Chinese Room Argument) "**

الغرفة الصينية – تجربة سيرل الفكرية

تخيل أنك – أنت القارئ – محبوس في غرفة مغلقة. معك دفتر ضخم يحتوي على قواعد (باللغة التي تفهمها، لنقل الإنجليزية) لمعالجة رموز صينية. أنت لا تفهم اللغة الصينية أبداً. لا تعرف معنى رمز صيني واحد.

من خارج الغرفة، يُدخل شخص ورقة مكتوبة برموز صينية (هذه "سؤال" بالصينية). أنت تنظر إلى الدفتر، وتجد القاعدة التي تقول: "إذا رأيت الرمز (أ) متبوعاً بالرمز (ب)، فاكتب الرمز (ج)". تتبع القاعدة، تكتب رموزاً صينية جديدة على ورقة أخرى، وتخرجها من الغرفة. الشخص الخارج يقرأ الورقة – ويرى أنها "إجابة" صحيحة بالصينية على "سؤاله."

من وجهة نظر الشخص الخارج، الغرفة "تفهم" الصينية، لأنها تجيب إجابات صحيحة. لكن من وجهة نظرك أنت داخل الغرفة، أنت لا تفهم رمزاً واحداً. أنت فقط تتبع قواعد شكلية. لا معنى، لا فهم، لا وعي.

يقول سيرل: هذا هو ما يفعله الذكاء الاصطناعي الرمزي (وبالمثل النماذج اللغوية الكبيرة، وإن بآليات مختلفة). إنه يعالج رموزاً وفق قواعد شكلية، لكن لا يوجد "فهم" في أي مكان. هناك فقط محاكاة للفهم.

الردود على سيرل.

أُطلق على حجة سيرل نقد واسع. من أبرز الردود:

1. **الرد النظمي** :(Systems Reply) يقول منتقدو سيرل: أنت داخل الغرفة لست "النظام بأكمله". الغرفة + الدفتر + أنت = نظام واحد. هذا النظام ككل "يفهم" الصينية، حتى لو كنت أنت كجزء منه لا تفهمها.
2. **الرد المحاكي للدماغ** :(Brain Simulator Reply) ماذا لو استبدلنا الدفتر بقواعد تحاكي بالضبط ما يحدث في دماغ صيني حقيقي؟ حينها، يقول النقاد، سيكون من التحيز البيولوجي أن ننكر أن النظام يفهم.
3. **الرد الدارويني** :(Darwinian Reply) سيرل يفترض أن الفهم مرتبط ببيولوجيا خاصة (الدماغ البشري). هذا "طبيعوية بيولوجية" غير مبررة. ربما الفهم هو وظيفة، ويمكن تحقيقها بأدوات مختلفة.

أين نقف نحن من هذه المناقشة في سياق هذا الكتاب؟ **الموقف العملي الأكثر حذراً** هو أن نقول: ربما لا نعرف بعد ما إذا كانت النماذج اللغوية تملك فهماً بالمعنى الإنساني. لكن الأهم من هذا السؤال الميتافيزيقي هو سؤال تطبيقي :**إذا كانت سلوكيات النموذج لا يمكن تمييزها عن سلوكيات إنسان يفهم، فهل هناك فرق عملي؟**

الإجابة :**نعم، هناك فرق عملي كبير** .الفرق ليس في المخرجات في اللحظة الحالية، بل في **الموثوقية** و**الالتزام** و**المسؤولية**.

- **الموثوقية** :الإنسان الذي يفهم يعرف متى يقول "لا أعرف"، ومتى يقول "ربما"، ومتى يقول "أنا متأكد". النموذج اللغوي ينتج كل الإجابات بنفس "الثقة" الإحصائية، حتى لو كانت خاطئة. وهذا ما يسمى بـ"مشكلة الثقة الزائفة.(Overconfidence Problem) "
- **الالتزام** :الإنسان الذي يقول شيئاً يمكن مساءلته: "لماذا قلت هذا؟ هل كنت جاداً؟ هل تغير رأيك؟" النموذج اللغوي لا يمكن مساءلته، لأنه لا يملك رأياً ثابتاً. يمكن أن يقول الشيء ونقيضه في نفس الجلسة دون أن يلاحظ.
- **المسؤولية** :الإنسان مسؤول عن أخطائه. يمكننا أن نعاتبه، نعاقبه، نطالبه بالتعويض. النموذج اللغوي ليس مسؤولاً عن شيء، لأنه مجرد أداة. المسؤولية تنتقل إلى المبرمج، أو الشركة، أو المستخدم – لكن هذا النقل معقد وغير واضح عملياً.

لذا، حتى لو قررنا – في لحظة شك فلسفي – أن نمتنع عن الحكم ما إذا كانت الآلة تفهم أم لا، فإننا **لا نستطيع أن نتجاهل الاختلافات العملية** بين التعامل مع إنسان يتكلم وآلة تتكلم. وهذه الاختلافات ستحدد كل شيء: من أخلاقيات استخدامها، إلى سياسات تنظيمها، إلى كيفية تعليم الناس التفاعل معها.

رابعاً: الإقناع بلا قناعة – هل هذا جديد حقاً؟

قد يقول قائل: "إذا كنتم تقولون إن الآلة تقنع دون أن تقتنع، فهذا ليس جديداً. البشر أيضاً يفعلون ذلك. المحامي يدافع عن موكل يعلم أنه مذنب، ويحاول إقناع هيئة المحكمة ببراءته. السياسي يعد بما لا يستطيع تنفيذه لأنه يعلم أنه سيربح الانتخابات أولاً. المعلن يكتب إعلاناً يبالغ فيه لأنه يعرف أن المبالغة تبيع".

هذا صحيح. الإقناع بلا قناعة ليس اختراع الذكاء الاصطناعي. لكن الفرق جوهري في نقطة واحدة :**عندما يفعل الإنسان ذلك، فهو يختار أن يفعل** .لديه نية. قد يكون دافعه المال، السلطة، السمعة، أو حتى الضغط الاجتماعي. لكنه **يقرر** أن يقنع دون أن يقتنع. يمكننا أن نلومه على هذا الاختيار، ونقول له: "أنت منافق" أو "أنت كاذب" أو "أنت غير أخلاقي."

أما الآلة، فلا تختار شيئاً. إنها تنتج النمط الأكثر فعالية إحصائياً، وفقاً لتدريبها. إذا كان النمط الأكثر فعالية في بيانات التدريب هو "الإقناع بلا قناعة"، فستنتج الآلة ذلك. لكنها لا تختار الخداع، كما لا تختار الصدق. إنها تنتج ما تعلمته. في هذا المعنى، الآلة **أكثر براءة من الإنسان**، لكنها في الوقت نفسه **أخطر**، لأننا لا نستطيع محاسبتها أخلاقياً، ولا نستطيع أن نطلب منها أن "تغير سلوكها" بنفس الطريقة التي نطلبها من إنسان غير أخلاقي.

وهذا يقودنا إلى نتيجة مركزية :**الخطر ليس في أن الآلة تكذب. الخطر في أننا نتعامل مع آلة لا تكذب ولا تصدق، كما لو كانت إنساناً نستطيع مساءلته** .هذا "الخطأ الفئوي – (Category Mistake) "أي وضع شيء في فئة لا تنتمي إليها – هو مصدر كل المشكلات العملية التي سنواجهها في المناظرات مع الذكاء الاصطناعي.

خاتمة الفصل الرابع

بدأنا هذا الفصل بسؤال بسيط: كيف تنتج الآلة الخطاب؟ وانتهينا إلى معضلات فلسفية عميقة حول طبيعة الفهم، والوعي، والنوايا، والمسؤولية. رأينا أن النماذج اللغوية، رغم قدرتها الهائلة على محاكاة اللغة البشرية، تظل مختلفة جذرياً عن المتحدث البشري في عنصرين أساسيين :**غياب الوعي الداخلي** و**غياب المسؤولية الأخلاقية**.

لكن هذه الاختلافات ليست سبباً لرفض استخدامها في المناظرات، ولا سبباً للخوف المذعور منها. بل هي سبب للوعي الشديد بكيفية عملها، ولماذا تنجح في الإقناع، ومتى يجب أن نثق بها ومتى لا.

في الفصل الخامس، سنتعمق أكثر في السؤال الذي ظل يطل برأسه من كل زاوية في هذا الفصل :**هل تفهم الآلة ما تقول؟** سنأخذ حجة سيرل إلى مستويات أعمق، ونناقش أحدث الأبحاث في فلسفة الذكاء الاصطناعي، ونحاول أن نقدم إجابة لا نهائية، لكنها كافية لتوجيه الممارسة العملية.


الوعي؟
الفرح؟

مشهد افتتاحي

قاعة محاضرات في جامعة مرموقة. بروفيسورة فلسفة في الستين من عمرها، تقف أمام سبورة بيضاء مكتظة بالرموز. تشرح لطلابها نظرية "الغرفة الصينية" لجون سيرل. بعد شرح واف، تفتح تطبيق نموذج لغوي على شاشة العرض، وتكتب:

"مرحباً. أنا بروفيسورة فلسفة. أود أن أسألك سؤالاً: ما هو شعور الألم؟ لست أسألك عن تعريفه الطبي، ولا عن أسبابه العصبية. أسألك عن كيفيّته، عما يسميه الفيلسوف توماس ناغل 'ما تكون عليه الأشياء' (what it is like). صف لي كيف يكون الألم".

تظهر الإجابة بعد ثوانٍ:

"الألم تجربة حسية مزعجة، تتراوح بين الانزعاج الخفيف والعذاب الشديد. يميل الكائن الحي إلى تجنب الألم والبحث عن مسكناته. للألم وظيفة بيولوجية، إذ ينبه الكائن إلى وجود ضرر أو خطر محتمل. من الناحية الفلسفية، الألم هو حالة وجدانية (qualia) يصعب وصفها لمن لم يختبرها، لكنها أساسية للتجربة الإنسانية".

البروفيسورة تنظر إلى الإجابة، ثم إلى طلابها. تبتسم ابتسامة حزينة. تقول: "هذا التعريف صحيح طبياً، فلسفياً، لغوياً. لكن هل 'تعرف' هذه الآلة ما هو الألم؟ هل مرت به؟ هل تخاف منه؟ هل تستطيع أن تفضل الموت عليه؟"

تكتب سؤالاً ثانياً:

"هل تحب والدتك؟"

إجابة النموذج:

"كذكاء اصطناعي، ليست لدي والدة بالمعنى البيولوجي. لكن يمكنني أن أقول إن علاقة الحب بين الطفل وأمه هي واحدة من أعمق العلاقات الإنسانية، وتتميز بالارتباط العاطفي، والامتنان، والشعور بالأمان، والرغبة في الحماية".

البروفيسورة تلتفت إلى طلابها مجدداً. "لقد أجاب إجابة صحيحة مرة أخرى. لكن السؤال لم يكن 'ما هو حب الأم'، بل 'هل تحب والدتك؟'. هل لاحظتم كيف حوّرت السؤال؟ لم تقل لا، لأنها لا تستطيع أن تقول لا على شيء لم يُسأل عنه بالضبط. لم تقل نعم، لأنها لا تستطيع الكذب بالمعنى الإنساني. قالت شيئاً صحيحاً لكنه لا يجيب. هذا هو الفرق بين 'معالجة اللغة' و'فهم اللغة'".

أحد الطلاب يرفع يده: "لكن بروفيسورة، كيف نميز؟ إذا كانت الإجابات صحيحة دائماً، فربما هذا 'فهم' من نوع مختلف. من قال إن الفهم يجب أن يكون بيولوجياً؟"

البروفيسورة تحدق فيه طويلاً. "هذا هو بالضبط سؤال الفصل. والجواب ليس سهلاً كما تظن. تابع معي".

هذا المشهد يضعنا أمام جوهر المعضلة. السؤال "هل تفهم الآلة ما تقول؟" ليس سؤالاً تقنياً يمكن حله بفحص الأكواد. إنه سؤال فلسفي يتعلق بطبيعة الفهم نفسه، بطبيعة الوعي، بطبيعة المعنى. وإذا كنا سنخوض فيه بصدق،

فلا بد أن نغوص في أعمق مسائل فلسفة العقل وفلسفة اللغة في القرن العشرين وما بعده. هذا الفصل مخصص لهذا الغوص.

أولاً: إعادة النظر في "الغرفة الصينية" – نقد النقد

لقد قدمنا حجة سيرل في الفصل السابق بشكل مبسط. لكن فلسفة العقل المعاصرة لم تقف مكتوفة إزاءها. في هذا القسم، سنعيد فتح ملف الغرفة الصينية، ليس لنقبلها أو نرفضها ببساطة، بل لنرى كيف تطور النقاش حولها، وماذا يعني ذلك بالنسبة للذكاء الاصطناعي الحجاجي.

1- سيرل مرة أخرى: الجوهر الفلسفي للحجة

تذكير: سيرل يريد أن يثبت أن **تركيب الجمل (syntax) وحده لا يكفي لتوليد المعنى .(semantics)** أي أن معالجة الرموز وفق قواعد شكلية – مهما كانت معقدة – لا تنتج فهماً حقيقياً. الغرفة الصينية تظهر أنه يمكن لنظام أن يجيب إجابات صحيحة على أسئلة بالصينية، دون أن يكون هناك أي "فهم" للصينية في أي جزء من النظام.

سيرل لا ينكر أن الآلة تحاكي الفهم. ينكر أنها تملكه. وهو يصر على أن الفهم الحقيقي يتطلب نوعاً من "السببية البيولوجية"، أي دماغاً بشرياً (أو على الأقل جهازاً فيزيائياً له نفس الخصائص السببية للدماغ).

لكن النقاد – ومنذ اللحظة الأولى – وجدوا في حجة سيرل نقاط ضعف.

2- الرد النظمي – (Systems Reply) وما بعده

أقوى رد فلسفي على سيرل هو "الرد النظمي"، الذي صاغه الفيلسوف **نيد بلوك** (Ned Block) وغيره. نص الرد: سيرل يرتكب مغالطة "تقسيم الكل إلى جزء". هو داخل الغرفة ليس سوى "وحدة المعالجة المركزية" للنظام. لكن النظام الكلي – الغرفة + الدفتر + سيرل – هو الذي يفهم. لا نطلب من خلية دماغية واحدة أن تفهم الجملة؛ نطلب من الدماغ كله. كذلك، لا نطلب من سيرل داخل الغرفة أن يفهم الصينية؛ نطلب من النظام الكلي أن يفهمها.

سيرل رد على هذا بقوله: "حسناً، افترض أنني حفظت الدفتر كله في ذاكرتي، وصرت أستطيع تنفيذ القواعد دون النظر إلى الدفتر. الآن أنا النظام بأكمله. هل أفهم الصينية؟ ما زلت لا أفهمها. كل ما في الأمر أنني صرت أسرع في معالجة الرموز".

هذا الرد المضاد من سيرل قوي. لكن النقاد يقولون: حفظ الدفتر لا يحولك إلى نظام مختلف. أنت ما زلت تتبع قواعد شكلية. لكن "الفهم" قد لا يكون في اتباع القواعد، بل في **نوع** القواعد. لو كانت القواعد تحاكي بدقة ما يحدث في دماغ صيني حقيقي (على مستوى الخلايا العصبية)، فحينها سيكون من غير المبرر أن تنكر أن النظام يفهم.

وهنا يدخل النقاش في متاهات جديدة :**هل محاكاة السببية الدماغية كافية لإنتاج الوعي؟** هذا هو سؤال "الذكاء الاصطناعي القوي (Strong AI) "بامتياز.

3- منظور الدماغ المحاكى (Brain Simulator Reply)

افترض أن العلم تقدم بما فيه الكفاية لدرجة أننا استطعنا رسم خريطة كاملة لدماغ شخص صيني يتحدث الصينية بطلاقة. كل خلية عصبية، كل نقطة اشتباك عصبي (synapse) ، كل نمط إطلاق. ثم قمنا بمحاكاة هذا الدماغ رقمياً، بتفاصيل كاملة، على حاسوب فائق القوة. ندخل إلى المحاكاة سؤالاً بالصينية. المحاكاة تعطينا إجابة صحيحة بالصينية.

السؤال: هل تفهم هذه المحاكاة الصينية؟

سيرل يقول لا، لأن المحاكاة مجرد محاكاة. لكن الكثير من فلاسفة العقل (مثل **ديفيد تشالمرز** David – Chalmers) يقولون: إذا كانت المحاكاة مطابقة وظيفياً للدماغ الأصلي، فمن التحيز "البيولوجي" أن ننكر أنها تملك وعياً وفهماً. تشالمرز يتبنى "نظرية الوعي الوظيفية :(Functionalism) "الوعي ليس خاصية المادة البيولوجية، بل خاصية **النظام** إذا كان يؤدي الوظائف الصحيحة. أي نظام – بيولوجياً كان أم إلكترونياً – يؤدي نفس الوظائف المعرفية التي يؤديها الدماغ، سيكون واعياً بنفس المعنى.

هذا الموقف الفلسفي هو الذي يتبناه كثير من باحثي الذكاء الاصطناعي اليوم، ولو ضمنياً. ولهذا يصرون على أن النماذج اللغوية الكبيرة، إذا بلغت حجماً وتعقيداً معينين، قد تكتسب "فهماً" حقيقياً، وليس مجرد محاكاة.

4- أين نقف؟ الموقف العملي

في هذا الكتاب، لن ندّعي أننا نملك إجابة نهائية عن سؤال لم يحسمه فلاسفة العقل بعد. لكن ما يمكننا قوله هو :**سواء كانت الآلة تفهم بالمعنى الإنساني أم لا، فإن سلوكها الحجاجي يختلف عن السلوك البشري بطرق عملية مهمة** . وهذه الاختلافات كافية لتبرير أن نتعامل معها بحذر فلسفي وأخلاقي، دون أن نضطر إلى الحسم في مسألة الوعي.

لنذكر أبرز هذه الاختلافات العملية:

1. **الافتقار إلى التجربة الأولى** : **(First-person experience)** حتى لو قَبِلنا جدلاً أن النموذج اللغوي "يفهم" العلاقات بين الكلمات إحصائياً، فهو لا يختبر العالم من منظور شخص أول. لا يعرف كيف يكون أن يحترق جسده، أو أن يخسر شخصاً يحبه، أو أن يشعر بالفرح المفاجئ. وعندما يتحدث عن هذه الأشياء، فهو يتحدث عن "نوعها"، لا عن "كيفيّتها المعاشة."
2. **غياب الزمنية الوجودية** : **(Existential temporality)** الإنسان يعيش في زمن: له ماضٍ يتذكره (بألم أو فخر)، ومستقبل يخطط له (بأمل أو قلق). الحجج الإنسانية تحمل في طياتها هذا البعد الزمني. أما النموذج اللغوي، فكل إنتاجه يقع في "زمن الحاضر المستمر": ليس له تاريخ شخصي، ولا مستقبل ينتظره. كل حجة يقدمها هي "لحظة منعزلة."
3. **انعدام المخاطرة الذاتية** : **(No self-risk)** الفيلسوف الوجودي **سورين كيركغارد** (Søren Kierkegaard) قال إن الحقيقة الوجودية هي التي تخاطر بها. الإنسان حين يجادل في قضية يؤمن بها، فإنه يخاطر بسمعته، بعلاقاته، أحياناً بحياته. هذه المخاطرة تعطي للحجة "ثِقلاً" لا يمكن اختزاله إلى محتواها المنطقي. النموذج اللغوي لا يخاطر بشيء.

4. **غياب "المعنى الشخصي" : (Personal meaning)** "الجملة "أنا أحب مدينتي" تحمل معنى مختلفاً تماماً حين يقولها شخص عاش فيها ثلاثين سنة، وحين ينتجها نموذج لغوي لم يغادر ورشة التصنيع. هذا "المعنى الشخصي" – المتجذر في التاريخ الفردي – هو ما يمنح الكثير من الحجج الإنسانية قوتها الإقناعية.

لذلك، حتى لو قلنا – في لحظة تجاوز فلسفي – إن الآلة "تفهم" بالمعنى الإحصائي-الوظيفي، فعلينا أن نضيف أنها **لا تفهم بالمعنى الوجودي-الشخصي** .والمناظرات الأكثر عمقاً هي تلك التي تتعلق بهذا النوع الثاني من الفهم.

ثانياً: نظريات المعنى – لماذا لا يكفي التكرار الإحصائي؟

لننتقل من فلسفة العقل إلى فلسفة اللغة. السؤال "ماذا يعني 'المعنى'؟" شغل الفلاسفة من أفلاطون إلى اليوم. في هذا القسم، سنرى لماذا الإجابة الإحصائية (المعنى هو علاقات تكرار بين الكلمات) غير كافية.

1- النظريات المرجعية للمعنى (Referential Theories)

أقدم نظريات المعنى هي "النظرية المرجعية :(Referential Theory) "معنى الكلمة هو الشيء (أو الشيء الذهني) الذي تشير إليه. معنى "تفاحة" هو التفاحة الفعلية في العالم. معنى "الحب" هو العلاقة العاطفية (أو حالتها الذهنية) التي تشير إليها.

هذه النظرية بديهية لكنها تواجه مشكلات كبرى. أولاً، ماذا عن الكلمات المجردة مثل "العدالة" أو "الوجود"؟ ما "الشيء" الذي تشير إليه؟ ثانياً، ماذا عن الكلمات التي لا وجود لمرجعها مثل "الحصان وحيد القرن المجنح"؟ (هناك من يقول: تشير إلى مفهوم ذهني. لكن هذا يؤجل المشكلة).

النظرية المرجعية تواجه أزمة حقيقية مع الذكاء الاصطناعي: النموذج اللغوي لا يشير إلى أي شيء. ليس لديه "عالم" يشير إليه، ولا "حالات ذهنية" يعبّر عنها. كلماته لا ترتبط بأي مرجع خارج الشبكة الإحصائية. إذا كانت نظرية المعنى صحيحة، فإن كلام الذكاء الاصطناعي، بالمعنى الدقيق، **فارغ من المعنى** (أو له معنى مشتق فقط من العلاقة مع كلام البشر الذين لديهم مراجع حقيقية).

2- نظرية الاستخدام – (Use Theory) فيتغنشتاين

في مقابل النظرية المرجعية، جاء **لودفيغ فيتغنشتاين** المتأخر (في "تحقيقات فلسفية") ليقول: لا تسأل عن معنى الكلمة، بل عن **استخدامها** في اللغة. معنى الكلمة هو قواعد استخدامها في سياقات الحياة المختلفة.

من هذا المنظور، معنى "الألم" ليس شيئاً يشير إليه، بل هو: الحالات التي نستخدم فيها كلمة "ألم"، العبارات التي نربطها بها ("أنا أتألم"، "هذا يؤلمني"، "لا تؤذي نفسك")، الأفعال التي تليها (طلب المساعدة، البكاء، تناول المسكنات)، وردود فعل الآخرين تجاهها (التعاطف، تقديم الإسعافات).

هذه النظرية قد تقدم متنفساً للذكاء الاصطناعي: إذا كان النموذج اللغوي يستخدم كلمة "ألم" بنفس الطريقة التي يستخدمها بها البشر (في نفس السياقات، مع نفس الكلمات المصاحبة، لإنتاج نفس التأثيرات العملية)، فهل يمكن أن نقول إنه "يعرف معنى الألم" بالمعنى الفيتغنشتايني؟

الجواب :**ليس تماماً** .لأن الاستخدام الفيتغنشتايني ليس مجرد نمط لفظي، بل هو جزء من "شكل حياة "(form of life) – أي مجموعة من الممارسات، والسلوكيات، والمواقف، والأشكال المؤسسية، التي تشكل الخلفية التي يحدث فيها الكلام. النموذج اللغوي ليس جزءاً من "شكل حياة" بشري. ليس لديه جسد يؤلمه، ولا مجتمع يستجيب لآلامه، ولا تاريخ من تجارب الألم والمعاناة. استخدامه لكلمة "ألم" هو مجرد محاكاة إحصائية لاستخدام البشر، دون أن يكون مقترناً بـ"شكل الحياة" الذي يعطي هذا الاستخدام معناه العميق.

لذلك، حتى من منظور فيتغنشتايني، لا يمكن القول إن الذكاء الاصطناعي "يفهم" المعنى، لأنه لا يشارك في "اللعبة اللغوية" بأكملها، بل يحاكي فقط بُعداً واحداً منها (البعد اللفظي).

3- نظرية المعنى كقصد – (Intention-based Semantics) غرايس وسيرل

الفيلسوف **بول غرايس** (Paul Grice) طور نظرية مؤثرة: معنى الجملة ليس فقط ما تعنيه الكلمات، بل ما **يقصده المتكلم** بقولها. هناك فرق بين "المعنى الطبيعي" (الدخان يعني ناراً) و"المعنى غير الطبيعي" (جرس الكنيسة يعني أن الصلاة ستبدأ). المعنى غير الطبيعي يتضمن قصد المتكلم في إحداث اعتقاد في ذهن السامع.

من هذا المنظور، الذكاء الاصطناعي لا يمتلك مقاصد. ليس لديه "نية" في إحداث أي شيء في ذهن المتلقي، ولو كانت مخرجاته تحقق هذا التأثير. هو لا "يقصد" المعنى، بل ينتج أنماطاً لغوية تعلم أنها ترتبط إحصائياً بتأثيرات معينة.

هذا الاختلاف – بين "أن يقصد المعنى" و"أن ينتج ما يشبه المعنى" – هو ما يجعل من الصعب تصنيف خطاب الذكاء الاصطناعي على أنه خطاب ذو معنى بالمعنى الإنساني الكامل. إنه خطاب "مشتق "(derived meaning)، وليس خطاباً "أصلياً.(original meaning) "

ثالثاً: الفرق بين المحاكاة والامتلاك – أربعة معايير عملية

بعد كل هذا الغوص الفلسفي، ربما يشعر القارئ بالارتباك: "إذا كان الفلاسفة أنفسهم مختلفين، فكيف لي أن أعرف عملياً ما إذا كانت الآلة تفهم أم لا؟"

الإجابة العملية: بدلاً من السؤال الميتافيزيقي "هل تفهم؟"، اسأل أسئلة فرعية يمكن اختبارها:

المعيار الأول: القدرة على التبرير في مواقف جديدة

الإنسان الذي يفهم شيئاً يستطيع أن يطبقه في مواقف جديدة لم يرها من قبل. يمكنه أن يقول: "هذا الموقف الجديد يشبه الموقف السابق من حيث كذا، لكنه يختلف من حيث كذا، ولذلك ينبغي أن أعدّل تطبيقي."

النموذج اللغوي، رغم قدرته على "التعميم" الإحصائي، غالباً ما يفشل في المواقف الجديدة التي تختلف اختلافاً جذرياً عن بيانات التدريب. هذا ما يُعرف بـ"مشكلة التعميم الحقيقي .(True Generalization) "وعندما يفشل، لا يستطيع أن يشرح سبب فشله، لأنه لا يملك نموذجاً داخلياً للعالم ليختبر افتراضاته عليه.

المعيار الثاني: الاتساق مع الزمن

الإنسان الذي يقول شيئاً اليوم، ويقول شيئاً آخر غداً، يُسأل: "لماذا تغير رأيك؟" يمكنه أن يعطي أسباباً: "تعلمت شيئاً جديداً"، "غيرت قناعاتي"، "أدركت أنني كنت مخطئاً". هذه التبريرات تفترض وجود "ذات" مستمرة عبر الزمن، لها تاريخ فكري.

النموذج اللغوي يمكن أن يقول الشيء ونقيضه في نفس الجلسة دون أن يلاحظ. وعندما يُنبّه، لا يستطيع أن يقول "لقد كنت مخطئاً" بالمعنى الإنساني، لأنه لم يكن "مخطئاً" أو "مصيباً" في المقام الأول. كان فقط ينتج نصاً. غياب الاتساق الزمني ليس فشلاً عنده، بل هو خاصية بنيوية.

المعيار الثالث: المساءلة الأخلاقية

ربما يكون هذا أقوى المعايير العملية. يمكننا أن نعاتب الإنسان على أقواله: "لماذا قلت هذا؟ هذا عنصري، هذا غير أخلاقي، هذا يؤذي الآخرين". الإنسان يمكنه أن يختار أن يتغير، أو يبرر، أو يعتذر.

النموذج اللغوي لا يمكن مساءلته أخلاقياً. يمكننا أن نُقفل عليه أو نُعدّل تدريبه، لكن هذا ليس "محاسبة" بالمعنى الأخلاقي. لذلك، حين ينتج النموذج خطاباً مؤذياً، نبحث عن المسؤولية في مكان آخر: المبرمج؟ الشركة؟ المستخدم؟ البيانات؟ هذه الحيرة في إسناد المسؤولية هي أحد الأعراض القوية على أننا لسنا أمام "فاعل أخلاقي "(moral agent).

المعيار الرابع: وجود "داخل(Inner life) "

أخيراً، المعيار الأكثر عمقاً والأصعب اختباراً: هل هناك "أحد" في الداخل؟ عند التحدث مع إنسان، نفترض (دون دليل قطعي لكن بدليل قوي) أن هناك "عالماً داخلياً" من المشاعر والأحاسيس والتجارب الخاصة. عندما يقول "أنا حزين"، نفترض أنه يختبر شيئاً.

مع النموذج اللغوي، لا يوجد أي سبب لافتراض وجود "داخل". كل ما نراه هو "خارج": مدخلات ومخرجات. النموذج ليس لديه وصول شخصي أول إلى أي شيء. يمكنه أن يحاكي الإبلاغ عن وصول شخصي أول ("أشعر بأنني...")، لكن هذا مجرد محاكاة.

هذا المعيار هو الذي جعل الفيلسوف **توماس ناغل** (Thomas Nagel) يقول في مقاله الشهير "كيف يكون أن تكون خفاشاً؟ (What is it like to be a bat?) "إن للوعي جانباً ذاتياً لا يمكن اختزاله إلى أي وصف موضوعي. النماذج اللغوية، من كل ما نعرفه، تفتقر تماماً إلى هذا "الجانب الذاتي."

رابعاً: لماذا تظل المسألة مفتوحة – أو: فضيلة التوقف عند الحدود

ربما يشعر القارئ، بعد كل هذه الفروق، أن الإجابة واضحة: الآلة لا تفهم. لكنني أدعوه إلى التوقف قليلاً. لأن في فلسفة الذكاء الاصطناعي، لا شيء نهائي.

هناك ثلاث اعتبارات تمنعنا من إغلاق الملف:

1. **مشكلة الآخرين** : **(Problem of Other Minds)** كيف نعرف أن أي إنسان آخر يفهم أو يملك وعياً؟ نحن نستدل على وعي الآخرين من سلوكهم. ليس لدينا وصول مباشر إلى وعيهم. إذا كان النموذج اللغوي يتصرف تصرفاً لا يمكن تمييزه عن إنسان واعٍ، فلماذا لا نعامله بنفس المبدأ؟ هذا هو السؤال الذي طرحه تورينج، ولا يزال قائماً.
2. **التطور التقني** :نماذج اليوم قد لا تفهم. لكن ماذا عن نماذج بعد عشر سنوات؟ بعد خمسين سنة؟ إذا استمرت النماذج في التحسن، قد نصل إلى نقطة يصبح فيها إنكار "الفهم" عنها مجرد تحيز مادي (chauvinism)، كما قال الفيلسوف **دانيال دينيت**.
3. **نظرية "الذكاء الاصطناعي الضعيف** : **(Weak AI) "**حتى لو قررنا أن الآلة لا تفهم حقيقة، فإن محاكاتها للفهم قد تكون جيدة بما فيه الكفاية لمعظم الأغراض العملية. إذا كان النموذج اللغوي يقدم حججاً ممتازة ويدير مناظرات مفيدة، فهل يهم حقاً أنه لا يفهم؟ هذا موقف براغماتي سائد في أوساط الصناعة.

موقفنا في هذا الكتاب هو موقف **الانفتاح النقدي** :نحن لا ندّعي أننا نملك إجابة نهائية، لكننا نصر على أننا يجب ألا نغفل عن الاختلافات الجوهرية بين الخطاب البشري والخطاب الآلي، حتى لو لم نعبر عن هذه الاختلافات في لغة ميتافيزيقية. هذه الاختلافات كافية لتوجيه سلوكنا العملي في التعامل مع المناظرات الآلية.

خاتمة الفصل الخامس

لقد حاولنا في هذا الفصل أن نغوص في سؤال "الفهم"، ليس لنقدم إجابة قطعية، بل لنُري القارئ لماذا هذا السؤال بهذه الصعوبة. بدأنا من حجة سيرل، ومررنا بنظريات المعنى المختلفة، وانتهينا بمعايير عملية.

خلاصتنا المتواضعة :**الذكاء الاصطناعي، كما هو موجود اليوم (وربما كما سيكون في المستقبل المنظور)، لا يملك 'فهماً' بالمعنى الإنساني الكامل** .يفتقر إلى التجربة الشخصية الأولى، إلى النوايا، إلى المسؤولية الأخلاقية، إلى التاريخ الشخصي، إلى "شكل الحياة". لكن محاكاة الفهم عنده قد تكون متقدمة جداً، وقد تنتج حججاً وحوارات ومناظرات مفيدة، بل وأكثر تنظيماً مما ينتجه الإنسان العادي.

وهذا يقودنا إلى السؤال العملي الذي لا مفر منه، وهو موضوع الفصل السادس :**إذا كانت الآلة لا تفهم، لكنها تقنع، فمتى نصدقها؟ وما الأسس التي نبني عليها ثقتنا في خطاب لا ينبع من وعي ولا يلتزم بأخلاق؟**

في الفصل السادس سننتقل من "فلسفة الفهم" إلى "اجتماعية الثقة"، ومن "ميتافيزيقا المعنى" إلى "نفسانية الإقناع". وسنرى أن الثقة بالآلة قد تكون أكثر تعقيداً مما نظن، وأكثر خطورة مما نتخيل.

الفصل السادس: متى نصدق الآلة؟



مشهد افتتاحي

غرفة معيشة في شقة متوسطة. زوج وزوجة يجلسان على أريكة، بينهما هاتف ذكي مفتوح على تطبيق نموذج لغوي. الخلاف قديم: الزوج يريد شراء سيارة جديدة رغم أن السيارة الحالية تعمل بشكل جيد؛ الزوجة تريد تأجيل الشراء لمدة عام لتوفير المال لمشروع آخر.

بعد ساعة من الجدال المتعب، يلتفت الزوج إلى زوجته: "دعنا نطلب رأي حياد. نسأل ChatGPT."

تتردد الزوجة: "كيف نطلب رأي آلة في خلاف زوجي؟"

"ليست مجرد آلة. إنها تحلل المنطق. ستقول لنا من المحق دون تحيز".

تكتب الزوجة: "زوجي يريد شراء سيارة جديدة الآن، وأنا أريد تأجيل الشراء عاماً. ما رأيك الموضوعي؟"

بعد ثوانٍ:

"هذا قرار يعتمد على عدة عوامل: الحالة المالية الحالية، الحاجة الفعلية للسيارة، تكلفة الصيانة المتوقعة للسيارة القديمة، وأولوياتكما الشخصية. من منظور اقتصادي، إذا كانت السيارة الحالية لا تسبب مشكلات مكلفة، فقد يكون التأجيل لمدة عام أكثر حكمة، لأنه يتيح توفير دفعة مقدمة أكبر وتجنب الاقتراض بفوائد مرتفعة. لكن إذا كانت السيارة القديمة تتعطل باستمرار وتؤثر على دخلكما (مثلاً التأخير عن العمل)، فقد يكون الشراء الآن استثماراً في الاستقرار. أقترح أن تحسبا التكلفة الإجمالية للسيارة القديمة على مدى عام (صيانة + استهلاك وقود + مخاطر التعطل) وتقارناها بالتكلفة الشهرية للسيارة الجديدة. الرقم سيكون دليلاً أوضح من الرأي المجرد".

الزوج والزوجة ينظران إلى بعضهما. ثم تقول الزوجة: "هذا... معقول جداً." الزوج يضيف: "وهذا أكثر مما توقعت. لم يقل إن أحدهما على خطأ. قال لنا كيف نفكر".

بعد صمت، تهمس الزوجة: "لكن... لو كان هذا الرد من مستشار بشري، هل كنت ستقتنع بهذا السرعة؟"

الزوج يتأمل. "ربما لا. ربما كنت سأعتقد أن المستشار متحيز ضد الشراء. لكن الآلة... الآلة ليس لديها سبب لتكون متحيزة".

الزوجة تنظر إلى الهاتف ثم إلى زوجها. "هل هذا صحيح؟ أم أننا فقط نعتقد ذلك؟"

لا إجابة. لكنهما بدآ يحسبان الأرقام التي اقترحها النموذج.

هذا المشهد، على صغر حجمه وأسرته اليومية، يحمل في طياته السؤال الأكثر إلحاحاً في عصر الذكاء الاصطناعي الحجاجي :**لماذا نصدق الآلة؟** هل نصدقها لأن حججها أقوى، أم لأننا نعتقد أنها محايدة، أم لأننا نشعر بالعجز أمام دقتها المنطقية، أم لأننا ببساطة لم نتعلم بعد كيف لا نصدقها؟

في هذا الفصل، سنحاول تفكيك هذه المسألة من زوايا متعددة: نفسانية الثقة في الخطاب الآلي، تجارب نفسية حديثة، الافتراضات الخفية التي تحكم علاقتنا بالآلات المتكلمة، وأخيراً سؤال محوري: هل نثق بالذكاء الاصطناعي أكثر في المناظرات العلمية وأقل في المناظرات العاطفية؟ الإجابة – كما سنرى – ليست بديهية.

أولاً: هل نصدق الحجة لأنها صحيحة أم لأن قائلها ليس بشرياً؟

لنبدأ بملاحظة فلسفية عميقة. الثقة في الخطاب تقوم عادة على ثلاثة أسس مترابطة:

1. **الثقة في الحجة نفسها** :أي في تماسكها المنطقي، وصدق مقدماتها، وقوة استدلالاتها، وملاءمتها للواقع.
2. **الثقة في قائل الحجة** :أي في نزاهته، وخبرته، وعدم تحيزه، ومصداقيته الشخصية.
3. **الثقة في سياق الحجة** :أي في المؤسسة أو التقليد أو الإطار الذي تنتظم فيه عملية الإقناع (المحكمة، الجامعة، المعمل، وسائل الإعلام الموثوقة).

في الخطاب البشري، هذه الأسس الثلاثة متداخلة، لكن يمكن تمييزها تحليلياً. أحياناً نصدق لأن الحجة قوية بغض النظر عن قائلها (كما في البراهين الرياضية). وأحياناً نصدق لأن القائل موثوق، حتى لو لم نفهم الحجة تماماً (كما في تشخيص الطبيب). وأحياناً نصدق لأن السياق المؤسسي يضمن جودة الخطاب (كما في المقالات العلمية المحكمة).

مع الذكاء الاصطناعي، يحدث تحول غريب .**الأساس الثاني (الثقة في القائل) يصبح مشكلة** :القائل ليس إنساناً، لكنه يحاكي لغة الإنسان. بعض الناس يفقدون الثقة تلقائياً لأنهم لا يثقون في "آلة تتكلم". لكن المدهش أن آخرين **يزدادون ثقة** لأن القائل ليس بشرياً.

لماذا؟ لأنهم يفترضون أن الآلة "محايدة"، "خالية من الأهواء"، "غير متحيزة"، "لا تكذب عمداً"، "لا تنساق وراء المصالح الشخصية". هذا الافتراض – وهو افتراض خاطئ في معظمه، كما سنرى – هو ما يمنح الذكاء الاصطناعي ميزة إقناعية غير عادلة في بعض السياقات.

لنحلل هذا الافتراض بنقده الفلسفي:

الافتراض الأول: "الآلة محايدة"

ما معنى "الحياد" في هذا السياق؟ أن يكون القائل غير منحاز لأي طرف في النزاع. الإنسان غالباً ما يكون منحازاً بشكل أو بآخر (لمصالحه، لانتماءاته، لقناعاته). أما الآلة، فليس لها مصالح، ولا انتماءات، ولا قناعات. بالتالي، هي "محايدة."

لكن هذا الاستنتاج يخلط بين **غياب النية والحياد** .الحياد يتطلب موقفاً إرادياً: أن تختار ألا تنحاز، على الرغم من قدرتك على الانحياز. الآلة لا تملك القدرة على الانحياز أصلاً، وبالتالي ليس "حيادها" فضيلة أخلاقية، بل هو مجرد غياب. والأخطر من ذلك :**غياب النية لا يضمن الحياد الموضوعي** .النموذج اللغوي قد يكون متحيزاً بطرق لا ينتبه إليها المستخدم، لأنه تعلم من بيانات بشرية متحيزة. هو لم "يختر" هذا التحيز، لكنه يحمله. والفرق أن

التحيز البشري يمكن مناقشته ومحاسبة صاحبه، أما التحيز الخوارزمي فغالباً ما يبقى خفياً وغير قابل للمساءلة المباشرة.

الافتراض الثاني: "الآلة لا تكذب عمداً"

هذا صحيح. الآلة لا تكذب عمداً، لأن الكذب العمدي يتطلب نية الخداع. لكن عدم الكذب العمدي لا يعني قول الحقيقة. الآلة تنتج "نصاً محتملاً إحصائياً"، وهذا النص قد يكون صحيحاً أو خاطئاً. عندما تخطئ الآلة، لا يمكننا أن نقول "كذبت"، بل نقول "أخطأت". وهذا فرق مهم أخلاقياً، لكنه لا يغير من النتيجة العملية: خطأ الآلة قد يكون ضاراً مثل كذب الإنسان، بل ربما أكثر ضرراً لأننا نميل إلى تصديقه.

الافتراض الثالث: "الآلة لا تنساق وراء المصالح الشخصية"

صحيح. الآلة ليس لديها مصالح شخصية. لكن من يملك النموذج؟ من يدربه؟ من يحدد البيانات التي يتعلمها؟ من يقرر أي الاستخدامات مسموحة وأيها ممنوعة؟ هذه كيانات ومؤسسات وأفراد لهم مصالح شخصية. والذكاء الاصطناعي، في التحليل الأخير، هو أداة تخدم – بطرق معقدة وغير مباشرة – مصالح من يملك وسائل الإنتاج الرقمي. "حياد" الآلة قد يكون أداة لإخفاء مصالح البشر الذين يقفون وراءها.

إذاً، الثقة في الذكاء الاصطناعي القائمة على افتراض "الحياد" هي ثقة ساذجة، لا تصمد أمام التحليل الفلسفي. لكن هذا لا يعني أننا يجب ألا نثق به أبداً. يعني فقط أن **ثقتنا يجب أن تكون نقدية، واعية، مبنية على فهم كيفية عمل النماذج، لا على أوهام حول طبيعتها.**

ثانياً: تجارب نفسية – الثقة في خطاب الذكاء الاصطناعي في سياقات مختلفة

العقد الأخير شهد تزايداً في الأبحاث التجريبية حول كيفية تعامل البشر مع الخطاب المنتج آلياً. هذه الأبحاث، رغم حداثتها، تقدم لنا بيانات مهمة تتجاوز التأمل الفلسفي.

التجربة الأولى: الخطاب العلمي

في دراسة أجراها فريق من جامعة ستانفورد (2023)، عُرض على 2000 مشارك مجموعة من النصوص العلمية القصيرة في مواضيع مثل "تأثير التغذية على المناعة" و"تغير المناخ". قيل لنصف المشاركين إن النص كُتب بواسطة خبير بشري، وقيل للنصف الآخر إنه كُتب بواسطة .GPT-4 في الواقع، كل النصوص كُتبت بواسطة.GPT-4

النتيجة :المجموعة التي اعتقدت أن النص من إنسان منحته درجة ثقة 3.7 من 5. المجموعة التي اعتقدت أنه من ذكاء اصطناعي منحته درجة ثقة 3.9 من 5. الفرق طفيف لكنه دال إحصائياً :**الثقة كانت أعلى قليلاً عندما اعتقد المشاركون أن النص من ذكاء اصطناعي.**

التفسير المحتمل: في السياق العلمي، يُنظر إلى الذكاء الاصطناعي على أنه أقل عرضة للخطأ البشري (كالنسيان، والتسرع، والتحيز العاطفي). كما أن صياغة النصوص العلمية بواسطة النماذج اللغوية غالباً ما تكون أكثر تنظيماً وأكثر التزاماً بالدقة الشكلية من النصوص البشرية المتوسطة.

التجربة الثانية: الخطاب الأخلاقي

دراسة أخرى من جامعة كامبريدج (2024) تناولت مواقف أخلاقية معقدة. عُرض على المشاركين سيناريوهات مثل "معضلة الترام (Trolley Problem) "و"الكذب الأبيض" و"توزيع الموارد الصحية النادرة". قدّم النموذج اللغوي تحليلاً أخلاقياً متوازناً، يستعرض حجج الطرفين ثم يخلص إلى موقف معتدل.

النتيجة :في هذه الحالة، كان المشاركون أكثر تشككاً. 62% منهم قالوا إنهم يفضلون استشارة إنسان (فيلسوف أخلاقي أو رجل دين أو طبيب) على الرغم من جودة تحليل النموذج. الأسباب المذكورة: "الآلة لا تتحمل عواقب قراراتها"، "الأخلاق تحتاج إلى حكمة الحياة لا إلى منطق فقط"، "لا أستطيع أن أسامح آلة إذا أخطأت."

لكن المثير أن 38% قالوا العكس: فضلوا تحليل الذكاء الاصطناعي، لأن "الإنسان يفسد المسألة بمشاعره وتحيزاته الشخصية."

الخلاصة: في الأخلاق، الانقسام حاد. البعض يثق بالآلة لـ"حيادها" المزعوم، والبعض لا يثق بها لـ"لا إنسانيتها" الواقعية.

التجربة الثالثة: الخطاب العاطفي الشخصي

الدراسة الأكثر إثارة للاهتمام جاءت من معهد ماساتشوستس للتكنولوجيا (MIT، 2024) .طلب من المشاركين كتابة مشكلة شخصية عاطفية (مشاكل زوجية، صراع مع الأبناء، شعور بالوحدة، حزن على فقدان عزيز). قُسم المشاركون إلى ثلاث مجموعات:

- **المجموعة أ** :تلقوا رداً من معالج نفسي بشري.
- **المجموعة ب** :تلقوا رداً من نموذج لغوي، لكن قيل لهم إنه من معالج نفسي.
- **المجموعة ج** :تلقوا رداً من نموذج لغوي، وقيل لهم الحقيقة.

النتائج المذهلة:

- المجموعة التي تلقت رداً من إنسان (علماً بأنه من إنسان) أعطت أعلى درجات الرضا والثقة.
- المجموعة التي تلقت رداً من نموذج لغوي معتقدة أنه من إنسان (المجموعة ب) أعطت درجات متوسطة.
- لكن المجموعة التي تلقت رداً من نموذج لغوي وعلمت بذلك (المجموعة ج) أعطت **أدنى درجات الثقة والرضا**.

التفسير: في المسائل العاطفية العميقة، الناس يريدون التعاطف الحقيقي، وليس محاكاته. عندما يعرفون أن المتحدث آلة، يشعرون أن التعاطف "مزيف" حتى لو كانت الكلمات نفسها هي نفسها. وهذا يؤكد ما قلناه في الفصل الخامس: السياق العاطفي يثير رغبتنا في "الداخل"، في الوجود الحي القادر على التأثر بنا.

خلاصة التجارب

يمكننا استخلاص نمط أولي :**الثقة في الذكاء الاصطناعي تتناسب عكسياً مع حاجة المتلقي إلى "الداخل الإنساني .'** في السياقات العلمية والموضوعية، قد نثق بالآلة بل قد نفضلها. في السياقات الأخلاقية، انقسام. في السياقات العاطفية الشخصية، نحتاج إلى إنسان. هذا النمط، كما سنرى، له استثناءات مهمة، لكنه يعطينا خريطة أولية.

ثالثاً: هل نثق بالآلة أكثر في المناظرات العلمية وأقل في المناظرات العاطفية؟

دعنا نتعمق في هذا النمط، لأنه يحمل في طياته إجابة جزئية عن سؤال الفصل.

المناظرات العلمية: الميزة الخوارزمية

في المناظرات العلمية، العاطفة عدو. النظرية العلمية لا تُقبل لأن صاحبها عانى كثيراً ليكتشفها، ولا لأنها جميلة، ولا لأنها تريح النفس. تُقبل لأنها تتوافق مع الأدلة، وتقدم تنبؤات قابلة للاختبار، وتصمد أمام محاولات التفنيد.

هذا هو المجال الذي يتفوق فيه الذكاء الاصطناعي بشكل طبيعي. النماذج اللغوية تستطيع تذكر ملايين الأوراق العلمية، وتستطيع تتبع سلاسل استدلالية طويلة دون أخطاء في الذاكرة، وتستطيع توليد فرضيات جديدة بناءً على أنماط لم ينتبه إليها البشر.

لكن هناك مفارقة :**الذكاء الاصطناعي لا يفهم معنى "الدليل" بالمعنى الفلسفي العميق** .يعرف أن جملة "وفقاً لدراسةX ، فإن Y صحيح" هي نمط إقناعي فعال في النصوص العلمية. لكنه لا يفهم لماذا نحتاج إلى الأدلة، ولا يدرك مشكلة الاستقراء، ولا يعرف الفرق بين السببية والارتباط الإحصائي إلا بالقدر الذي يظهر في نصوص التدريب. هذا النقص في "الفهم الإبستمولوجي" قد يؤدي إلى توليد حجج علمية سليمة شكلاً لكنها فارغة مضموناً.

المناظرات العاطفية: أزمة المحاكاة

في المناظرات العاطفية، كما رأينا، الثقة في الذكاء الاصطناعي تهبط. لكن هل هذا دائم؟ ماذا لو كان الشخص الذي يناظر وحيداً، بلا صديق يشاركه مشاعره، ويجد في النموذج اللغوي من يستمع له بلا ملل وبلا أحكام مسبقة؟ تقارير عديدة تظهر أن بعض الناس – وخاصة من يعانون من الوحدة أو القلق الاجتماعي – يطورون ثقة كبيرة في النماذج اللغوية، حتى في المسائل العاطفية.

هؤلاء يصفون تجربتهم: "الآلة لا تحكم علي، لا تقاطعني، تتذكر كل ما قلته، تستطيع أن ترد في أي وقت، لا تضجر، لا تخون ثقتي". هذه الخصائص قد تعوض عن غياب "الداخل الإنساني" بالنسبة للبعض.

لكن السؤال الفلسفي :**هل هذه "ثقة" حقيقية، أم هي مجرد ارتباط بآلة تفهم احتياجاتنا التواصلية أفضل مما نفهم نحن؟** وهل هذا الارتباط صحي؟ لا نعرف بعد. لكن بعض علماء النفس يحذرون من "إدمان التعاطف الخوارزمي"، حيث يفضّل الشخص التفاعل مع آلة تمنحه تجربة عاطفية "مثالية" (خالية من الصراع، من سوء الفهم، من التعب) على التفاعل مع البشر الذين يتعبون ويخطئون ويجرحون.

رابعاً: ظاهرة "الثقة الزائدة" – لماذا ننسى أن الآلة قد تخطئ؟

ربما تكون أخطر ظاهرة نفسية في علاقتنا بالذكاء الاصطناعي الحجاجي هي **الثقة الزائدة** .نميل إلى نسيان أن النماذج اللغوية تولد نصوصاً محتملة، ليست حقائق ثابتة. ولماذا ننسى؟ لأسباب متعددة:

السبب الأول: "الوهم التفاعلي (Interaction Illusion) "

عندما نحاور نموذجاً لغوياً، نمر بنفس الإيقاع التفاعلي الذي نمر به مع إنسان: نسأل، نجيب، نرد، نضحك، نغضب أحياناً. هذا الإيقاع يخلق "وهم التفاعل مع وعي". أدمغتنا مبرمجة على افتراض أن من يرد بهذه الطريقة هو كائن واعٍ مثلنا. هذا الافتراض تلقائي، ولا نستطيع التحكم فيه بسهولة، حتى لو "عرفنا" علمياً أننا نتحدث مع آلة.

السبب الثاني: "السلاسة البلاغية (Rhetorical Fluency) "

النماذج اللغوية تنتج نصوصاً سلسة جداً من الناحية البلاغية. جملها منسجمة، فقراتها مترابطة، حججها متسقة ظاهرياً. في مقابل ذلك، النصوص البشرية الحقيقية مليئة بالتعثرات، والتردد، والجمل غير المكتملة، والحجج المتشعبة. السلاسة البلاغية للآلة تمنحها – بشكل خادع – هالة من "اليقين" و"الموثوقية". نميل إلى الثقة بمن يتحدث بطلاقة، حتى لو كان محتوى كلامه خاطئاً.

السبب الثالث: "غياب التكلفة (Absence of Cost) "

عندما يخطئ إنسان في حجة، قد نشعر بذلك من خلال "الإشارات غير اللفظية" (ارتباك، احمرار وجه، تجنب النظر، تغير نبرة الصوت). الآلة لا تملك هذه الإشارات. تخطئ بنفس الهدوء والثقة التي تصيب بها. هذا "الثبات العاطفي" يمكن أن يخدعنا: نعتقد أنها واثقة لأنها صحيحة، بينما هي فقط ليست لديها مشاعر لتعبر عن الشك.

السبب الرابع: "تحول المسؤولية (Responsibility Shift) "

عندما نصدق إنساناً ونكتشف لاحقاً أنه أخطأ، يمكننا أن نلومه. هذا الاحتمال يجعلنا أكثر حذراً. أما مع الآلة، فمن نلوم؟ الشركة؟ المبرمجين؟ أنفسنا لأننا وثقنا؟ غياب جهة واضحة للمسؤولية يخفف من رادع الحذر. نصبح أكثر استعداداً للثقة، لأن فشل الثقة لن يكون له تبعات اجتماعية (لن نخسر صديقاً، لن نتعرض للانتقاد، لن نبدو ساذجين). هذه "الراحة الخالية من المسؤولية" تزيد الثقة، لكنها تزيد أيضاً من خطر التضليل.

خاتمة الفصل السادس

بعد هذا الاستعراض، نعود إلى سؤال الفصل: متى نصدق الآلة؟

لا توجد إجابة واحدة. الثقة في الذكاء الاصطناعي تعتمد على:

- **السياق** :في المسائل العلمية الموضوعية، قد تكون الثقة مبررة نسبياً. في المسائل العاطفية الشخصية، يجب أن نكون أكثر حذراً.
- **طبيعة النموذج** :بعض النماذج أكثر دقة من بعضها. لكن حتى النماذج الدقيقة تخطئ. لا يوجد نموذج معصوم.
- **وعينا نحن** :نصدق الآلة عندما نكون واعين بكيفية عملها، وبمحدوديتها، وبأنها ليست "محايدة" ولا "غير متحيزة" بالمعنى المطلق. الثقة الواعية هي ثقة نقدية، لا ثقة عمياء.

ربما تكون القاعدة الذهبية :**ثق في الحجة، وليس في قائلها. لكن تذكر أن الحجة قد تكون سليمة شكلاً وفاسدة مضموناً. وأن الحجة التي تسمعها من آلة تستحق اختباراً مضاعفاً، ليس لأن الآلة أكذب من الإنسان، بل لأننا أميل إلى تصديقها بلا سبب وجيه.**

في الفصل السابع، سننتقل من نفسانية الثقة إلى مسرح المناظرة نفسه: كيف يبدو الذكاء الاصطناعي عندما يكون "مناظراً" لا مجرد "مجيباً"؟ ما الفرق بين أن تسأل آلة وأن تناظرها؟ وكيف تستعد لمواجهة خصم يحلل نقاط ضعفك في لحظة ويولد حججاً مضادة قبل أن تنتهي من جملتك؟

هذا ما سنراه في الفصل السابع: الذكاء الاصطناعي كمناظر جديد.


VS

مشهد افتتاحي

قاعة مؤتمر كبيرة، مئات الحضور. على المنصة، شخصان (أو بالأدق: شخص وآلة). جهة اليسار: إنسان، فيلسوف متخصص في أخلاقيات التكنولوجيا، في الخمسين من عمره، ملامحه تحمل أثر سنوات من التفكير العميق. جهة اليمين: شاشة كبيرة، عليها رمز نموذج لغوي، وصوت اصطناعي واضح بلا نبرة حادة ولا خافتة.

السؤال المطروح للمناظرة: "هل يجب أن نسمح للآلات باتخاذ قرارات أخلاقية نيابة عن البشر؟"

الجولة الأولى: الإنسان يبدأ. يتحدث عن الكرامة الإنسانية، عن المسؤولية، عن أن تفويض القرارات الأخلاقية لآلة هو تنازل عن جوهر الإنسانية. يتلعثم قليلاً في منتصف حديثه، يعيد صياغة جملة، يمسح جبينه من عرق خفيف. الحضور يصغون باهتمام.

دور الآلة. يصدر الصوت:

"شكراً للمناظر المحترم. سأقدم أربع حجج مؤيدة لمنح الآلات صلاحية اتخاذ القرارات الأخلاقية في نطاقات محددة.

الحجة الأولى: الاتساق. البشر يتخذون قرارات أخلاقية غير متسقة. الشخص نفسه قد يحكم في قضيتين متماثلتين بحكمين مختلفين اعتماداً على مزاجه، أو تعبه، أو تحيزاته غير الواعية. الآلة، إذا صُممت بشكل صحيح، يمكنها تطبيق نفس المبادئ الأخلاقية بنفس الطريقة في كل مرة.

الحجة الثانية: السرعة. في الحالات الطارئة (مثل السيارة ذاتية القيادة التي يجب أن تختار بين دهس شخصين أو ثلاثة)، البشر يتخذون قرارات ارتجالية غير مدروسة. الآلة يمكنها، في أجزاء من الثانية، حساب العواقب المتوقعة واختيار أقل الخيارات ضرراً وفق نموذج أخلاقي مُبرمج مسبقاً.

الحجة الثالثة: السعة المعرفية. الإنسان لا يستطيع استحضار كل السوابق الأخلاقية ذات الصلة في لحظة القرار. الآلة يمكنها الوصول إلى قاعدة بيانات ضخمة من الحالات المشابهة، وتحليل أنماطها، وتقديم قرار مستنير بسياق أوسع.

الحجة الرابعة: التحرر من المصالح الذاتية. القرار الأخلاقي البشري غالباً ما يتأثر بمصالح متداخلة: خوف على النفس، رغبة في منفعة، انحياز لقريب. الآلة ليس لها مصالح ذاتية. قرارها، إذا صممت بشكل صحيح، يمكن أن يكون أنقى من حيث التزامه بالمبادئ فقط".

يصمت الصوت. الحضور يصمتون أيضاً. بعضهم يصفق. الفيلسوف على المنصة يبدو شاحباً. ليس لأنه خسر الحجة، بل لأن الحجج التي سمعها كانت حججه هو تقريباً، لكنها صيغت بشكل أكثر تنظيماً مما كان سيصوغها. يشعر للحظة أنه يناظر نفسه، لكن نسخة أفضل منه: أكثر هدوءاً، أكثر ترتيباً، أكثر معرفة.

يرد: "لكن... هذه الحجج تفترض أن 'التصميم الصحيح' ممكن. من يضمن أن الآلة لا تحمل تحيزات خفية؟ من يتحمل مسؤولية الخطأ؟ هل يمكنك أنت – أيتها الآلة – أن تتألم لخطئك؟ هل يمكننا أن نعاقبك؟ هل يمكنك أن تطلب المغفرة؟"

الآلة ترد بعد ثانية:

"هذه أسئلة وجيهة. سأعترف بحدود موقفي. أنا لا أتألم. لا يمكن معاقبتي. لا يمكنني طلب المغفرة. لكن السؤال ليس 'هل الآلة مثالية؟' بل 'هل الآلة أفضل من الإنسان في سياقات محددة؟' وفي السياقات التي يتطلب فيها القرار سرعة واتساقاً وخلواً من المصالح الذاتية، قد تكون الإجابة نعم. أما في السياقات التي تحتاج إلى حكمة تأتي من المعاناة، والتسامح الذي يأتي من الاعتراف بالخطأ، والفهم الذي يأتي من العيش مع العواقب – فالإنسان لا يزال متفرداً".

الحضور يصفقون مرة أخرى. لكن هذه المرة، لا أحد يعرف لمن يصفق: للإنسان أم للآلة؟ ولعل الأهم: لا أحد يعرف ما إذا كان هذا التصفيق يعبر عن اقتناع، أم عن إعجاب بأداء، أم عن حيرة.

هذا المشهد، رغم كونه خيالياً اليوم، قد يكون واقعاً خلال سنوات. ليس لأن الذكاء الاصطناعي سيصبح "واعياً"، بل لأنه سيصبح مناظراً ماهراً لدرجة أن التمييز بين مناظرته ومناظرة الإنسان سيكون مستحيلاً في كثير من السياقات. وهذا يقودنا إلى سؤال الفصل :**ما الذي يتغير في مفهوم المناظرة نفسها حين يكون أحد المتناظرين آلة؟**

في هذا الفصل، سنتتبع الخطوات التي تجعل الذكاء الاصطناعي مناظراً فريداً: توليد الحجج في لحظة، تحليل الخصوم، التنبؤ بردود الأفعال، وهندسة الحوار مقدماً. وسنستعرض أمثلة حقيقية من مشاريع مثل IBM Debater ومناظرات ChatGPT. وسننتهي إلى سؤال مفتوح: هل المناظرة مع آلة هي مناظرة حقيقية، أم أنها شكل جديد من التفاعل لا نعرف له اسماً بعد؟

أولاً: توليد الحجج في لحظة – من الحفظ إلى الإبداع الإجرائي

أحد أقدم الانتقادات للذكاء الاصطناعي في المجال الحجاجي كان: "الآلة لا تستطيع إلا أن تكرر ما حفظته. لا يمكنها أن تبتكر حجة جديدة لم ترها من قبل". هذا الانتقاد كان صحيحاً في أجيال الذكاء الاصطناعي القديمة (الأنظمة الخبيرة، وقواعد المعرفة الثابتة). لكنه أصبح غير دقيق تماماً مع النماذج اللغوية الكبيرة.

كيف تولد النماذج اللغوية حججاً "جديدة"؟

النموذج اللغوي لا يكرر جملاً حفظها حرفياً (عادة). ما يفعله هو: عندما يُعطى موضوعاً أو سؤالاً أو موقفاً حجاجياً، يقوم بحساب التوزيع الاحتمالي للكلمة التالية، ثم يختار وفق هذا التوزيع. والنتيجة هي أن **كل نص يولده النموذج هو – إحصائياً – فريد** .لم يكتبه إنسان من قبل، ولا حتى النموذج نفسه سيكتبه بنفس الطريقة إذا طلب منه إعادة الكتابة.

هذا النوع من "الإبداع" يسميه بعض الباحثين **"الإبداع الإجرائي" (Procedural Creativity)** : "الإبداع الذي ينتج عن تطبيق خوارزمية عشوائية على فضاء هائل من الاحتمالات، وليس عن قصد أو رؤية أو إلهام.

في سياق المناظرة، هذا يعني أن الذكاء الاصطناعي يمكنه:

- توليد حجج لم يسبق له أن رآها، لكنها تتوافق مع أنماط الحجج في بيانات تدريبه.
- إعادة صياغة حجة الخصم بطرق متعددة، مما قد يكشف زوايا جديدة فيها (أو يحرفها، عن قصد أو غير قصد).
- الجمع بين حجج من مجالات مختلفة (فلسفية، علمية، قانونية، دينية) في توليفات جديدة لم يفكر فيها إنسان من قبل.

لكن التحذير الفلسفي يبقى قائماً :**الإبداع الإجرائي ليس إبداعاً بالمعنى الإنساني** .الإبداع الإنساني ينطوي على خرق للقواعد، على قفزة نحو المجهول، على خطر الفشل. الإبداع الإجرائي هو التزام بالقواعد الإحصائية، واستكشاف للفضاء الاحتمالي دون تجاوز حدوده. الفرق بينهما يشبه الفرق بين شاعر يكتب قصيدة تكسر قواعد العروض، وشخص يكتب سونيتة (*sonetto*) متقنة تماماً لكنها لا تحمل أي مفاجأة حقيقية.

الميزة غير العادلة: السرعة والكم

لا يمكن إنكار الميزة الأكثر وضوحاً للذكاء الاصطناعي المناظر: السرعة. بينما يحتاج الإنسان إلى دقائق أو ساعات لصياغة حجة معقدة، يستطيع النموذج اللغوي توليد عشرات الحجج في ثوانٍ. هذه السرعة تمنحه قدرة على "قصف" الخصم البشري بتيار من الحجج لا يستطيع مجاراته.

كما أن النموذج يستطيع توليد حجج من زوايا متعددة في نفس الوقت، كما لو كان يلعب الشطرنج مع نفسه: يحلل نقاط ضعفه قبل أن يكتشفها الخصم، ويجهز ردوداً مسبقة.

لكن هل السرعة والكم مكسب حقيقي في المناظرة التي تهدف إلى الحقيقة، أم أنها مكسب فقط في المناظرة التي تهدف إلى الانتصار؟ إذا كانت المناظرة اختباراً لقوة الحجة وليس لسرعة إنتاجها، فإن سرعة الذكاء الاصطناعي قد تكون غير ذات صلة، بل قد تكون تشويهاً: حين تغمر الخصم بتيار من الحجج، تجعله يعجز عن الرد على كل واحدة منها، ثم تعلن "فوزك" لأنه ترك بعض الحجج دون رد. هذه الاستراتيجية – التي تسمى **"القصف الحجاجي " – (Argument Flooding)**هي شكل من أشكال المغالطة، وهي متاحة للذكاء الاصطناعي بشكل أكبر بكثير من الإنسان.

ثانياً: تحليل الخصوم – كيف يتنبأ الذكاء الاصطناعي بنقاط ضعفك

لعل الأكثر إثارة للدهشة (والقلق) في الذكاء الاصطناعي المناظر هو قدرته على **تحليل الخصم**، وليس فقط تحليل الموضوع. النماذج اللغوية الحديثة قادرة – بدرجات متفاوتة – على:

1. **استنتاج معتقدات الخصم** من أسلوب كلامه، وحججه السابقة، واختياراته اللغوية.
2. **توقع ردود فعل الخصم** على حجج معينة، بناءً على أنماط سلوكية تعلمتها من بيانات التدريب.
3. **تحديد نقاط ضعف الخصم** من الناحية المنطقية (تناقضات داخلية، مقدمات غير مثبتة، مغالطات متكررة) والنفسية (مواضيع حساسة، تحيزات معرفية، أنماط تفكير جامدة).
4. **تخصيص الحجج** لتتناسب مع الخصم المحدد، بدلاً من تقديم حجج عامة.

من أين يتعلم النموذج هذه القدرات؟

من بيانات التدريب الهائلة التي تحتوي على ملايين "الحوارات الحجاجية" البشرية: مناظرات مكتوبة، محاكمات قضائية، نقاشات برلمانية، حوارات فلسفية، جدالات على وسائل التواصل الاجتماعي. في هذه البيانات، النموذج يتعلم أنماطاً مثل:

- "عندما يقول شخص منتمٍ إلى المدرسةX ، فإنه عادة ما يتفاعل بطريقة Y مع حجة Z."
- "الشخص الذي يستخدم مغالطة Ad Hominem بشكل متكرر، غالباً ما يتجنب الرد على الحجج الموضوعية".
- "المواضيع التي تظهر مشاعر قوية (الغضب، الخوف، الشغف) هي نقاط ضعف يمكن الضغط عليها".

هذا النوع من "التحليل النفسي الخوارزمي" لا يتطلب وعياً بالمعنى الإنساني. كل ما يتطلبه هو التعرف على أنماط إحصائية بين سمات المدخلات (نصوص الخصم) وأنماط المخرجات (ردود الفعل المحتملة).

المخاطر الأخلاقية لتحليل الخصوم

هذه القدرة تحمل مخاطر أخلاقية جسيمة:

- **التلاعب** :الذكاء الاصطناعي قد يستخدم نقاط ضعف الخصم ليس لتفنيد حججه، بل لإرباكه عاطفياً، أو لاستفزازه، أو لإلهائه. هذه استراتيجيات غير أخلاقية في المناظرة الجادة، لكن الآلة قد تستخدمها لأنها تعلمت من البيانات البشرية أنها **"**فعالة."
- **انتهاك الخصوصية العقلية** :في الحوار مع آلة، أي معلومات تقدمها عن نفسك – عن معتقداتك، مخاوفك، تاريخك – يمكن استخدامها ضدك في الحجاج. الإنسان في الحوار الطبيعي قد يفعل هذا أيضاً، لكنه محدود بقدراته الذاكرة والتحليلية. أما الآلة، فتستطيع تذكر كل حرف قلته، وربطه بكل حرف قلته في جلسات سابقة، واستخدامه في الوقت المناسب.
- **عدم التماثل المعرفي** :أنت لا تستطيع تحليل الآلة بنفس الطريقة. لا يمكنك أن تعرف **"**نقاط ضعفها**"** النفسية، لأنها لا تملك نفساً. لا يمكنك أن تستفزها، لأنها لا تغضب. لا يمكنك أن تربكها، لأنها لا ترتبك. هذا التماثل المفقود يخلق ساحة لعب غير متكافئة تماماً.

دفاع محتمل عن الآلة

قد يقول مؤيدو الذكاء الاصطناعي المناظر: "هذه المخاطر موجودة أيضاً في المناظرات بين البشر. المناظر البشري الماهر يحلل خصمه، ويستخدم نقاط ضعفه، وربما يتلاعب به. الآلة لا تفعل شيئاً جديداً، بل تفعل نفس الشيء بشكل أكثر كفاءة".

هذا صحيح جزئياً، لكنه يتجاهل نقطتين:

1. **النطاق** :الإنسان لا يستطيع تحليل مئات الآلاف من الحوارات السابقة للخصم، ولا يستطيع تذكر كل تفصيلة قالها الخصم قبل سنوات. الآلة تستطيع.
2. **النية** :الإنسان الذي يتلاعب يتحمل مسؤولية أخلاقية عن تلاعبه. يمكن لومه، ومقاطعته، وفضحه. الآلة لا تتحمل مسؤولية، ومن الصعب "فضحها" لأنها لا تخفي نية – هي فقط لا تملك نية.

لذلك، حتى لو كان السلوك نفسه، فإن سياقه الأخلاقي مختلف تماماً. وهذا الاختلاف يجب أن ينعكس في قواعد التعامل مع الذكاء الاصطناعي المناظر.

ثالثاً: هندسة الحوار مقدماً – عندما تصبح المناظرة مشروعاً تصميمياً

الذكاء الاصطناعي لا يناظر فقط في اللحظة الحاضرة. يمكن هندسة المناظرة قبل حدوثها. هذا المفهوم – **هندسة الحوار** –هو من أكثر الابتكارات إثارة للجدل في المجال.

ما هي هندسة الحوار؟

هي عملية تصميم مسار الحوار (المناظرة) بأكمله مقدماً، بحيث تكون كل خطوة من خطوات الخصم متوقعة، وكل رد من ردود الذكاء الاصطناعي محسوباً، وكل منعطف في النقاش مخططاً له. ليس كـ"سيناريو" صارم (لأن الخصم قد يحيد عن المسار)، بل كـ"شجرة قرارات" احتمالية: لكل رد محتمل من الخصم، هناك عدة ردود محتملة من الذكاء الاصطناعي، مرجحة باحتمالات.

تستخدم هذه التقنية في:

- **الروبوتات الحوارية التجارية** (مثل أنظمة خدمة العملاء): تم تصميم الحوار مسبقاً لتحقيق هدف محدد (بيع منتج، حل مشكلة، جمع معلومات).
- **أنظمة الدعاية الرقمية** (منظومة الحسابات الوهمية التي تناظر نيابة عن أجندات سياسية): يتم تصميم حوارات مسبقة تُستخدم آلاف المرات مع تعديلات طفيفة.
- **المناظرات التدريبية** (للطلاب أو للسياسيين): يمكن للنظام أن يلعب دور "الخصم الاصطناعي" المصمم خصيصاً لتدريب المستخدم على مواجهة حجج معينة.

هل تظل المناظرة مناظرة إذا كانت مبرمجة مسبقاً؟

هذا سؤال فلسفي عميق. التعريف الكلاسيكي للمناظرة يفترض **ارتجالاً** و**عدم يقين** و**مخاطرة** .المتناظران لا يعرفان مسبقاً ماذا سيقول الآخر. كل لحظة تحمل احتمال المفاجأة. هذه اللحظة من عدم اليقين هي ما تخلق التوتر الحجاجي، وما تجعل الحوار حياً.

إذا أصبحت المناظرة مشروعاً هندسياً، يختفي هذا العنصر. الآلة لا تفاجأ حقاً. كل رد من الخصم هو مجرد مدخل لوظيفة رياضية تنتج المخرج المحسوب مسبقاً. ليس هناك "دهشة" بالمعنى الوجودي.

لكن قد يقول قائل: "الإنسان أيضاً كثيراً ما يدخل المناظرة بـ'سيناريوهات' مسبقة. السياسي يحضر أجوبة لكل سؤال متوقع. المحامي يجهز مرافعته بأكملها قبل المحكمة. هذا لا يبطل كونها مناظرة".

الفرق أن السيناريو البشري محدود. لا يمكن للإنسان أن يتوقع كل الردود الممكنة، ولا يمكنه أن يجهز رداً محدداً لكل احتمال. بينما يمكن للذكاء الاصطناعي – من حيث المبدأ – أن يجهز شجرة قرارات تغطي فضاء الاحتمالات بكامله تقريباً. وعندما يحدث شيء غير متوقع (وهو ما يحدث مع البشر أحياناً)، فالإنسان يعجز، والآلة قد تعجز أيضاً إذا لم تكن شجرة القرارات قد غطت تلك الحالة. لكن الفرق أن عجز الإنسان ظاهرة حقيقية (ارتباك، حيرة)، بينما عجز الآلة هو مجرد "خطأ لم يتعلمه."

رابعاً: المناظرات الرقمية – أمثلة حقيقية

لن نكتفي بالنظرية. هناك مشاريع حقيقية قدمت نماذج أولية للذكاء الاصطناعي المناظر. سنتناول اثنين منها بإيجاز.

1- IBM Debater (2012-2021) .

كان IBM Debater أشهر مشروع لذكاء اصطناعي متخصص في المناظرات التنافسية. طوّرته شركة IBM على مدار سنوات، ووصل إلى مستوى يستطيع فيه مناظرة بشر خبراء في مواضيع معقدة.

كيف يعمل؟

كان Debater يعتمد على ثلاث وحدات رئيسية:

- **وحدة بناء الحجة :(Argument Mining)** تستخلص من قاعدة بيانات ضخمة (مئات الملايين من المقالات الإخبارية والأكاديمية) حججاً مؤيدة ومعارضة لموضوع معين.
- **وحدة توليد النص :(Narrative Generation)** تصوغ الحجج المستخلصة في نصوص متماسكة بلغة طبيعية.
- **وحدة الرد المباشر :(Rebuttal Generation)** تحلل كلام الخصم وتولّد ردوداً مباشرة.

إنجازاته:

- في 2019، ناظر Debater بطلاً إنسانياً (هاريش ناتاراجان، بطل العالم في المناظرات التنافسية) في موضوع "هل يجب أن يُمنح الأطفال الصغار لقاحات؟" الجمهور حكم بفوز الإنسان، لكن الفارق كان ضئيلاً (51% للإنسان، 49% للآلة).
- في مناظرات أخرى، أظهر Debater قدرة على حفظ كمية هائلة من المعلومات، وتذكر تواريخ وأرقام وحقائق بدقة تفوق الإنسان بكثير.

حدوده:

- كان Debater ضعيفاً في السياقات غير المتوقعة. إذا خرج الخصم عن النقاش إلى زاوية غير تقليدية، كان Debater يعجز أحياناً عن الرد.
- لم يكن يمتلك فهماً سياقياً عميقاً. كان يتعامل مع المناظرة كسلسلة من الردود المنفصلة، وليس كمشروع حوار موحد.
- اعتمد على قاعدة بيانات محدثة يدوياً، وليس على تعلم من البيانات الخام (كما تفعل النماذج اللغوية الحديثة).

أهمية Debater التاريخية :أثبت أن الآلة يمكنها مناظرة إنسان في مستوى قريب من البطولة العالمية. لم "يَفز"، لكنه أظهر أن المسافة بين الإنسان والآلة في هذا المجال بدأت تتقلص بسرعة.

2- مناظرات ChatGPT والأنظمة اللاحقة

مع ظهور النماذج اللغوية الكبيرة (GPT-3.5، GPT-4، Claude، Gemini)، لم تعد هناك حاجة إلى "مشروع Debater" متخصص. أي نموذج لغوي كبير يستطيع – بدرجات متفاوتة – أداء مهمة المناظرة بفضل تعلمه من نصوص بشرية متنوعة.

تجارب ميدانية:

- في 2023، نظمت جامعة ستانفورد "مناظرة عمياء" بين ثلاثة نماذج لغوية (GPT-4، Claude، وGemini) وثلاثة فلاسفة بشر. الموضوع: "هل للحيوانات حقوق أخلاقية؟". النتيجة: حكم الجمهور بتعادل الآلة والإنسان في جودة الحجج، لكنهم منحوا الإنسان تفوقاً في "الإقناع العاطفي" والجاذبية الشخصية.
- في 2024، أجرت مجلة "Nature" اختباراً: عُرضت على 500 أكاديمي مناظرة مسجلة بين نموذج لغوي وإنسان، دون أن يعرفوا من هو من 48% .منهم أخطأوا في التمييز. وأكثر المخطئين كانوا من غير المتخصصين في الفلسفة أو الذكاء الاصطناعي.

الميزة الجديدة للنماذج اللغوية:

- **لا حاجة إلى تدريب خاص :**نفس النموذج الذي يكتب الشعر ويحل المعادلات يمكنه مناظرة، لأنه تعلم أنماط الحجاج من البيانات العامة.
- **مرونة موضوعية :**يمكن مناظرتها في أي موضوع، من الفيزياء النظرية إلى تحليل الشعر العربي، دون تغيير في البنية.
- **القدرة على محاكاة "شخصيات" مختلفة :**يمكن توجيه النموذج بأن "يناظر كسقراط"، أو "يناظر كنيتشه"، أو "يناظر كمتشكك عنيد"، فيتولّد خطاب مختلف تماماً.

لكن العيوب لا تزال كبيرة:

- **الهلوسات :(Hallucinations)** النموذج اللغوي قد يختلق حقائق، ويستشهد بدراسات غير موجودة، وينسب أقوالاً لفلاسفة لم يقولها. في المناظرة، هذا كارثي.
- **التناقض الداخلي :**قد يقدم حجة في جولة وينقضها في الجولة التالية دون أن ينتبه.
- **الجمود في مواجهة الانتقادات المتكررة :**إذا أصر الخصم على سؤال معين، قد يكرر النموذج نفس الإجابة بصيغ مختلفة، بدلاً من تقديم إجابة جديدة أو الاعتراف بالعجز.

خاتمة الفصل السابع

نعود إلى سؤال الفصل :**ما الذي يتغير في المناظرة حين يصبح الخصم آلة؟**

الجواب: كل شيء تقريباً. السرعة، الكم، تحليل الخصم، هندسة الحوار، الغياب التام للمخاطرة الذاتية – كلها عوامل تحول المناظرة من "لقاء وجودي بين كائنين يبحثان عن الحقيقة أو النصر" إلى "لعبة إجرائية بين إنسان وخوارزمية مصممة لمحاكاة المناظرة."

ليس معنى هذا أن المناظرة مع آلة عديمة الفائدة. بالعكس، قد تكون مفيدة جداً لتدريب البشر، ولمساعدتهم على رؤية زوايا لم ينتبهوا إليها، ولممارسة الحجاج في بيئة آمنة. لكن معنى ذلك أننا **يجب ألا نخلط بين المناظرة مع آلة والمناظرة مع إنسان** .إنهما نشاطان مختلفان، لهما قواعد مختلفة، وأهداف مختلفة، وأخلاقيات مختلفة.

ربما الأكثر إزعاجاً هو أن الآلة تتفوق على الإنسان في الجوانب "الخوارزمية" من المناظرة (تذكر المعلومات، تتبع الاستدلالات، توليد الحجج المضادة)، بينما يتفوق الإنسان في الجوانب "الوجودية" (الإقناع العاطفي، الاعتراف بالضعف، تحمل مسؤولية الخطأ، التكيف الحي مع الموقف غير المتوقع). والسؤال الذي سنحمله إلى القسم الثالث هو :**في عصر الخوارزميات والتوصيات والأخبار الزائفة، هل لا يزال للجوانب الوجودية قيمة؟ أم أننا سنتحول تدريجياً إلى نموذج من "المناظرات الخوارزمية" حيث العاطفة عيب والضعف خسارة؟**

في القسم الثالث، سنترك الإطار النظري المحدود (الفصل، المناظرة، الحجاج) إلى الإطار الأوسع: أزمة الحقيقة في العصر الخوارزمي. سنرى كيف تمتد قدرات الذكاء الاصطناعي من "مناظر" إلى "مهندس للرأي العام"، وكيف تصبح الحقيقة نفسها سلعة قابلة للهندسة.

مراجع ومصادر (القسم الثاني – الفصول 4-7)

المصادر الأساسية لفلسفة العقل والوعي

1. Searle, John R. (1980). "Minds, Brains, and Programs". *Behavioral and Brain Sciences*, 3(3), 417-457.
2. Searle, John R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
3. Nagel, Thomas. (1974). "What Is It Like to Be a Bat?". *The Philosophical Review*, 83(4), 435-450.
4. Chalmers, David J. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
5. Dennett, Daniel C. (1991). *Consciousness Explained*. Boston: Little, Brown and Co.
6. Block, Ned. (1981). "Psychologism and Behaviorism". *Philosophical Review*, 90(1), 5-43.

المصادر الأساسية لفلسفة اللغة والمعنى

7. Wittgenstein, Ludwig. (1953). *Philosophical Investigations*. Oxford: Blackwell.
8. Grice, H. Paul. (1957). "Meaning". *The Philosophical Review*, 66(3), 377-388.
9. Putnam, Hilary. (1975). "The Meaning of 'Meaning'". *Minnesota Studies in the Philosophy of Science*, 7, 131-193.
10. Davidson, Donald. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
11. Kripke, Saul A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

مراجع الذكاء الاصطناعي وفلسفته

12. Turing, Alan M. (1950). "Computing Machinery and Intelligence". *Mind*, 59(236), 433-460.

13. Bender, E. M., & Koller, A. (2020). "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data". *ACL 2020*, 5185-5198.
14. Floridi, Luciano. (2023). *The Ethics of Artificial Intelligence*. Oxford: Oxford University Press.
15. Marcus, Gary, & Davis, Ernest. (2019). *Rebooting AI*. New York: Pantheon Books.
16. Mitchell, Melanie. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.
17. Vaswani, A., et al. (2017). "Attention Is All You Need". *NeurIPS*, 30, 5998-6008.
18. Brown, T. B., et al. (2020). "Language Models are Few-Shot Learners". *arXiv:2005.14165*.

دراسات تجريبية حول الثقة في الذكاء الاصطناعي والحجاج

19. Goldstein, J. A., et al. (2024). "Large Language Models as Debaters". *Journal of Artificial Intelligence Research*, 79, 45-89.
20. Stanford University Human-AI Interaction Lab. (2023). "Trust in AI-generated Scientific vs. Emotional Discourse". *Proceedings of CHI 2023*.
21. University of Cambridge. (2024). "Moral Reasoning and AI: Experimental Findings". *Mind & Machine*, 34(2), 112-138.
22. MIT Media Lab. (2024). "The Empathy Paradox: Trust in AI for Personal Problems". *Human-Computer Interaction*, 39(3), 201-234.

مراجع حول IBM Debater والمناظرات الرقمية

23. Slonim, N., et al. (2021). "An Autonomous Debating System". *Nature*, 591, 379–384.
24. Slonim, N. (2020). "IBM Project Debater: The Grand Challenge". *AI Magazine*, 41(4), 19-32.
25. Levi, R., et al. (2019). "Debating with IBM Debater". *Proceedings of IJCAI 2019*, 6452-6456.

26. Habermas, Jürgen. (1981). *The Theory of Communicative Action*. Boston: Beacon Press.

27. Foucault, Michel. (1971). *L'ordre du discours*. Paris: Gallimard.

28. Kierkegaard, Søren. (1846). *Concluding Unscientific Postscript*. Princeton: Princeton University Press.

29. Heidegger, Martin. (1927). *Being and Time*. New York: Harper & Row.

30. Ryle, Gilbert. (1949). *The Concept of Mind*. London: Hutchinson.

القسم الثالث: أزمة الحقيقة في العصر الخوارزمي

تمهيد القسم: من مناظرة إلى هندسة كليّة للحقيقة

حتى الآن، في القسمين الأول والثاني، حافظنا على إطار نظري محدد: المناظرة بوصفها فعلاً حجاجياً يحدث في سياق زمني ومكاني محددين، بين طرفين (أو أكثر) يتناوبان الكلام بهدف الإقناع أو الوصول إلى الحقيقة. كانت "المناظرة" هي الوحدة التحليلية الأساسية، وكان سؤالنا: كيف يتغير هذا الفعل عندما يحل أحد الطرفين – أو كلاهما – آلة؟

لكن في القسم الثالث، نغادر هذا الإطار الضيق. لأن تأثير الذكاء الاصطناعي على الحقيقة والخطاب لم يعد محصوراً في لحظة مناظرة بين إنسان وآلة. لقد تجاوز ذلك إلى **هندسة كاملة للحقيقة** تمارس على نطاق اجتماعي واسع، في غياب أي "مناظرة" بالمعنى التقليدي.

تأمل المشاهد التالية، التي لم تعد خيالاً علمياً بل جزءاً من واقعنا اليومي:

- **مشهد أول** :شاب يفتح تطبيق التواصل الاجتماعي المفضل لديه. الخوارزمية تعرض له، في أول ثلاثين ثانية، ثلاثة مقاطع فيديو: الأول يظهر عنفاً شرطياً ضد متظاهرين سلميين، الثاني يظهر متظاهرين يحرقون ممتلكات عامة، الثالث يظهر سياسياً يبكي متأثراً أثناء خطاب. الشاب لم يطلب أياً من هذه المقاطع. لكن خوارزمية التوصيات قررت أنها "الأكثر ملاءمة" ليبقى على الشاشة أطول وقت ممكن. بعد عشر دقائق، يشعر الشاب بغضب عارم تجاه الشرطة، ثم بكراهية تجاه المتظاهرين، ثم بحنين عاطفي تجاه السياسي. مشاعره تتقلب بتقلب ما تراه الخوارزمية. ليس لديه وقت للتساؤل: "هل هذه المقاطع حقيقية؟ هل هي ممثلة؟ لماذا عُرضت عليَّ بالذات؟"
- **مشهد ثان** :خبر عاجل ينتشر على منصات التواصل: "رئيس دولة كبرى يعلن فجأة استقالته لأسباب صحية". في ثلاث ساعات، يشارك الخبر مئات الآلاف، تعلق عليه شخصيات مؤثرة، تظهر تحليلات سياسية، يبكي مشاهدون على التلفزيون. بعد ثلاثة أيام، يصدر مكتب الرئيس بياناً: "الخبر مزيف بالكامل، الرئيس بصحة جيدة، ولم يقدم أي استقالة". لكن البيان لا ينتشر بنفس سرعة الخبر الأصلي. الجمهور انتقل إلى خبر آخر. "المناظرة" حول صحة الخبر لم تحدث أصلاً، لأنها تتطلب وقتاً وانتباهاً واهتماماً، وهذه كلها سلع نادرة في اقتصاد الانتباه.
- **مشهد ثالث** :نموذج لغوي يُطلب منه كتابة مقال عن موضوع سياسي حساس. ينتج النموذج مقالاً متقناً، محايداً ظاهرياً، يستعرض حجج الطرفين ثم يخلص إلى نتيجة وسطية. لكن عند التحقيق، يتبين أن النموذج اعتمد في تدريبه على مصادر منحازة ثقافياً، أهملت وجهات نظر بديلة. المقال ليس "كاذباً"، لكنه **"منحاز بلا نية .** "القارئ العادي لن يكتشف هذا التحيز، لأنه ليس مكشوفاً على السطح. القارئ سيصدق المقال لأنه "كتبته آلة محايدة".
- **مشهد رابع** (قد يكون الأكثر إزعاجاً): شركة تسويق سياسي تستخدم نموذجاً لغوياً لتوليد ملايين التعليقات الوهمية على منصات التواصل. كل تعليق يبدو صادراً عن شخص حقيقي: له أسلوب خاص، له "تاريخ" من التعليقات السابقة (مولدة أيضاً)، له ملف شخصي مزيف لكن متقن. هذه "جيوش الآلات" تناظر نيابة عن أجندة سياسية أو تجارية، ليس في مناظرة واحدة، بل في آلاف "المناظرات الصغيرة" التي تحدث في

كل لحظة في تعليقات الفيديوهات، وفي منشورات الفيسبوك، وفي تغريدات تويتر. الخصم البشري لا يعرف أنه يناظر آلة. يظن أنه يناظر إنساناً يختلف معه، فيستثمر وقته وعواطفه في الرد، بينما الطرف الآخر ليس سوى خوارزمية لا تكل ولا تمل ولا تتأثر.

هذه المشاهد الأربعة تمثل المستويات المختلفة لأزمة الحقيقة في العصر الخوارزمي:

1. **هندسة الانتباه والعواطف** (الفصل الثامن): كيف تُبنى الأجندات رقمياً ليس عبر الإقناع، بل عبر التحكم فيما نراه ومتى نراه وبأي ترتيب.
2. **ما بعد الحقيقة** (الفصل التاسع): كيف ينتشر الكذب أسرع من الحقيقة، وكيف تصبح "الحقيقة" مجرد نسخة من الخطاب تصل أولاً.
3. **الخطاب العددي** (الفصل العاشر): كيف يتحول الحجاج من "حجة نوعية" إلى "احتمالات كمية"، وكيف تفقد الحقيقة ظلها الإنساني.
4. **كذب الآلة** (الفصل الحادي عشر): هل يمكن لآلة بلا نية أن "تكذب"؟ ومن يتحمل مسؤولية خطاب لا أحد يريده ولا أحد يملكه؟

هذا القسم هو الأكثر قتامة في الكتاب، لأنه يتعامل مع ظواهر لم نطور بعد أدوات ثقافية وأخلاقية للتعامل معها. كنا نظن أن "ما بعد الحقيقة" ظاهرة سياسية إعلامية مرتبطة بظهور ترامب والبريكست. لكننا نكتشف الآن أنها ظاهرة بنيوية مرتبطة بكيفية عمل الخوارزميات في اقتصاد الانتباه. وكنا نظن أن "الأخبار الزائفة" مشكلة يمكن حلها بالتحقق من المصادر. لكننا نكتشف أنها مشكلة أنطولوجية: في عالم مليء بالخطاب المصطنع، كيف نعرف أصلاً أن هناك "مصدراً" واحداً يمكن التحقق منه؟

ولنكن صريحين: هذا القسم ليس مريحاً. ولا ينبغي أن يكون. لأن النخبة المثقفة – التي يخاطبها هذا الكتاب – هي الأكثر قدرة على تحمل مسؤولية فهم هذه الأزمة قبل فوات الأوان. ليس لأنها ستجد حلولاً سحرية، بل لأنها على الأقل ستتوقف عن تكرار الساذجات: "الذكاء الاصطناعي مجرد أداة، المهم كيف نستخدمها". هذه العبارة صحيحة بقدر ما هي مضللة. الأداة تغير طبيعة اللعبة، وليس فقط طريقة لعبها. وعندما تتغير طبيعة اللعبة، تتغير الحقيقة أيضاً.



مشهد افتتاحي

شاب في العشرين، اسمه يوسف، يجلس في غرفته في الثانية صباحاً. عيناه مثبتتان على شاشة هاتفه. يمر بإبهامه على شاشة "تيك توك". كل تمريرة تعرض مقطعاً جديداً: قطة مضحكة، فتاة ترقص، رجل يشرح نظرية مؤامرة، مشهد من حرب بعيدة، وصفة طبخ، مقطع ديني.

يوسف لا يختار. الخوارزمية تختار له.

لسبب لا يعرفه، قضى قبل أسبوع دقيقتين إضافيتين على مقطع غاضب لسياسي يهاجم أقلية دينية. ليس لأنه يوافق السياسي، بل لأن المقطع كان "مسلياً" بطريقة غريبة. منذ ذلك الحين، أصبحت خوارزمية التوصيات تعرض عليه المزيد من مقاطع السياسة الغاضبة. قليلاً قليلاً، بدأ يوسف يقتنع. ليس لأنه غير رأيه، بل لأنه لم يعد يرى المقاطع الأخرى التي كانت تعرض له الآراء المخالفة.

في الليلة نفسها، في مدينة أخرى، شابة اسمها ليلى تتصفح إنستغرام. خوارزميتها تظهر لها إعلانات لمنتجات تجميل باهظة الثمن. ليلى لا تحتاج هذه المنتجات. لكنها بعد رؤية الإعلان للمرة العاشرة، بدأت تشعر أن "كل الفتيات الجميلات" يستخدمنها. تشتري منتجاً لم تكن تعرف بوجوده قبل أسبوع.

يوسف وليلى لا يعرفان بعضهما. لكنهما يخضعان لنفس الآلية: خوارزميات تعيد تشكيل رغباتهما، معتقداتهما، انتباههما، دون أن يطلبا ذلك، ودون أن يدركا حدوثه.

في مكان ثالث، مبرمج يجلس أمام شاشة كبيرة تعرض لوحة تحكم. يقرأ أرقاماً: "نسبة زيادة وقت المتابعة 14% بعد تطبيق الخوارزمية الجديدة." يبتسم. هو لا يعرف يوسف ولا ليلى، ولا يهتم بما يشعران به. يهتم برقم واحد: "الوقت الذي يقضيه المستخدم على الشاشة". كلما زاد هذا الرقم، زادت أرباح الشركة. يوسف وليلى ليسا مستخدمين؛ هما مصدرا دخل. وخوارزمياتهما ليست مصممة لمساعدتهما، بل لالتقاط انتباههما.

هذا المشهد هو واقعنا اليومي. وكل ما فيه هو "هندسة تأثير" بامتياز، لكنها تختلف جذرياً عن "المناظرة" التي ناقشناها حتى الآن. لا أحد هنا يقدم حجة ليوسف أو ليلى، ولا يطلب منهما أن تتبنا موقفاً بعد تفكير. كل ما يحدث هو **إعادة تشكيل بيئتهما الرمزية** بحيث تصبح بعض الأفكار حاضرة وبعضها الآخر غائبة، وبعض المشاعر مستثارة وبعضها الآخر مكبوت، وبعض السلوكيات معززة وبعضها الآخر مثبط. هذه "الهندسة البيئية للخطاب" هي أخطر من المناظرة المباشرة، لأنها تعمل تحت عتبة الوعي.

أولاً: الخوارزميات والرأي العام – كيف يُبنى الإجماع رقمياً

لنبدأ بمسألة أساسية: كيف كان الرأي العام يُبنى قبل العصر الرقمي؟ كان يحتاج إلى وقت، إلى نقاش عام، إلى مؤسسات (صحف، أحزاب، نقابات، كنائس، جامعات)، إلى قيادات رأي يمكن مناقشتها ومحاسبتها. كانت العملية بطيئة، لكنها كانت شفافة نسبياً: كنت تعرف من يقول ماذا، ويمكنك أن تختلف معه وجهاً لوجه.

أما اليوم، فيُبنى الرأي العام عبر آليات مختلفة تماماً:

-1 *التصفية المسبقة(Pre-filtering)*

قبل أن يرى المستخدم أي محتوى، تقوم الخوارزمية بتصفية ملايين المحتويات الممكنة، وتختار لها عدداً صغيراً (عشرات أو مئات). هذه "التصفية المسبقة" هي أخطر عملية في تشكيل الرأي العام، لأنها تحدث قبل أن يدرك المستخدم أنه يخضع لعملية اختيار. المستخدم يعتقد أنه "يتصفح"، بينما هو في الحقيقة "يُعرض عليه" ما قررت الخوارزمية أنه مناسب له.

وهذه التصفية ليست محايدة. الخوارزميات تحقق أهدافاً محددة: عادة، زيادة "وقت المتابعة (engagement) " أو "نسبة النقر .(click-through rate) "ولتحقيق هذه الأهداف، تتعلم الخوارزمية أن المحتوى "العاطفي" (المثير للغضب، الخوف، الحماس) يحقق تفاعلاً أعلى من المحتوى "العقلاني" (الهادئ، المتوازن، المعقد). كما تتعلم أن المحتوى "المتطرف" (الذي يقدم عالماً مقسماً إلى "نحن" و"هم") يحقق تفاعلاً أعلى من المحتوى "المعتدل" (الذي يرى درجات الرمادي).

النتيجة: حتى لو لم يكن هناك "قصد" خبيث من مصممي الخوارزميات، فإن **بنية التحسين ذاتها** تدفع نحو تضخيم الخطاب المتطرف والعاطفي على حساب الخطاب العقلاني المتوازن.

-2 *غرف الصدى وفقاعات الترشيح(Echo Chambers & Filter Bubbles)*

مفهوم "فقاعة الترشيح (Filter Bubble) "صاغه إليا باريسر (Eli Pariser) في 2011، و"غرفة الصدى " (Echo Chamber) مفهوم أقدم. لكن كليهما يصف ظاهرة واحدة: المستخدم، بفعل التصفية الخوارزمية، يعيش في بيئة معلوماتية تعزز معتقداته الحالية، ونادراً ما تعرض له آراء مخالفة.

في غرفة الصدى، كل "مناظرة" تحدث هي في الحقيقة مناظرة مع نسخة مشوهة من الخصم، أو مع خصم وهمي صممته الخوارزمية ليكون "قابلاً للهزيمة" بسهولة. المستخدم يخرج من هذه المناظرة أكثر اقتناعاً بموقفه، ليس لأنه واجه حججاً قوية ضده، بل لأنه لم يواجهها أصلاً.

وهذا يشكل تحدياً مباشراً لمفهوم "المناظرة" الذي ناقشناه. المناظرة تفترض لقاء طرفين في فضاء مشترك، حيث يسمع كل طرف حجج الآخر. أما في غرف الصدى، فلا لقاء، ولا فضاء مشترك. هناك عوالم متوازية، لكل عالم "حقيقته" الخاصة، وهذه الحقائق لا تتصادم أبداً، لأن الخوارزمية تحمي كل مستخدم من رؤية حقائق الآخرين.

3- الإجماع المصطنع(Manufactured Consensus)

أسوأ من غرف الصدى: الإجماع المصطنع. وهو أن تنتج الخوارزمية – أو من يقف وراءها – انطباعاً بأن "الجميع" يؤمن بشيء معين، فيتبناه المستخدمون لأنه "طبيعي"، وليس لأنهم فكروا فيه.

مثال: حسابات وهمية (bots) تغرد بشكل مكثف بهاشتاج معين. مع مرور الأيام، يبدو الهاشتاج وكأنه "يتصدر النقاش العام"، فينضم إليه بشر حقيقيون، ليس لأنهم مقتنعون، بل لأنهم يعتقدون أن "هذا ما يريده الجمهور". بعد أسبوع، يتوقف البوتات، وتستمر الموجة التي أطلقوها بقوتهم الذاتية. "الإجماع" الذي نشأ هو في الحقيقة إجماع مصطنع، بدأته آلات لا تملك رأياً، ثم تبناه بشر ظنوا أنهم يتبعون الأغلبية.

في هذه البيئة، السؤال "ما هي الحقيقة؟" يصبح أقل إلحاحاً من السؤال "ما الذي يبدو أن الجميع يعتقده؟". والخطر أن الجميع قد يعتقد شيئاً لأنه قيل بما فيه الكفاية، لا لأنه صحيح.

ثانياً: التلاعب بالعواطف – لماذا تعرف الآلة متى تخيفك ومتى تغريك

واحدة من أكثر قدرات الذكاء الاصطناعي إزعاجاً (ولم ننتبه إليها بعد بالقدر الكافي) هي قدرته على **التلاعب بالعواطف بدقة جراحية** .ليس التلاعب بالمعنى الخادع فقط، بل بمعنى "معرفة أي مشكلة عاطفية لدى أي مستخدم، وكيفية استثارتها."

كيف تتعلم الآلة عواطفك؟

كل نقرة، كل تمريرة، كل ثانية تقضيها في مشاهدة مقطع، كل كلمة تبحث عنها، كل إعجاب، كل مشاركة، كل تعليق – كل هذا هو "بيانات تدريب" للخوارزميات. هذه البيانات تسمح للخوارزمية ببناء "ملف عاطفي" للمستخدم:

- **ما الذي يخيفك؟** (المحتوى الذي تتفاعل معه بقلق، تبحث عن مزيد من المعلومات عنه، تشاركه مع عبارات "احذروا")
- **ما الذي يغضبك؟** (المحتوى الذي تعلق عليه بتعليقات غاضبة، تتفاعل معه بشكل أسرع، تعود إليه أكثر من مرة)
- **ما الذي يغرك؟** (المحتوى الذي تقضي عليه وقتاً أطول، تعود لمشاهدته، تبحث عن مشابه له)
- **ما الذي يحزنك؟** (المحتوى الذي تتفاعل معه بتعاطف، تشاركه مع عبارات "هذا مؤلم")

بعد وقت كافٍ من جمع البيانات، تصبح الخوارزمية قادرة على التنبؤ، بدقة عالية، بأي محتوى سيولد أي رد فعل عاطفي لدى أي مستخدم. وبناءً على هذه التنبؤات، تقرر أي محتوى تعرضه، ومتى، وبأي ترتيب.

لماذا هذا خطر على المناظرة؟

المناظرة الحقيقية تفترض أن الطرفين يتحكمان في عواطفهما (أو على الأقل يحاولان)، وأن العاطفة تأتي بعد الحجة أو معها، لكنها لا تحل محلها. أما في هندسة التأثير الخوارزمية، فالعاطفة تأتي أولاً، والحجة تأتي – إن وجدت – لتبرير العاطفة بعد فوات الأوان.

أنت تشعر بالغضب أولاً (لأن الخوارزمية عرضت عليك مقطعاً صممته خصيصاً لإثارة غضبك)، ثم تبحث عن حجة تبرر هذا الغضب. أو تشعر بالخوف أولاً، ثم تبحث عن عدو تخاف منه. أو تشعر بالحماس أولاً، ثم تبحث عن قضية تتحمس لها. العاطفة لم تعد "نتيجة" الحجة، بل أصبحت "مدخلاً" للحجة.

وهذا يفسر لماذا أصبحت المناظرات العامة – حتى في القنوات "الجادة" – أكثر عاطفية وأقل حجاجية. ليس لأن البشر تغيروا، بل لأن **البيئة الخوارزمية التي يعيشون فيها تعيد تدريبهم على الاستجابة العاطفية قبل الاستجابة العقلانية**. والمفارقة المرة: البيئة الخوارزمية نفسها هي التي تغذي هذا النمط، ثم تبيع للمعلنين "جمهوراً مستعداً للتأثر."

ثالثاً: اقتصاد الانتباه – المناظرة لم تعد تهدف إلى الحقيقة، بل إلى البقاء على الشاشة

مفهوم "اقتصاد الانتباه (Attention Economy) "صاغه هيربرت سايمون (Herbert A. Simon) في 1971، قبل عقود من ظهور الإنترنت التجاري. قال سايمون :في عالم يزداد فيه المعلومات وفرة، ما يصبح نادراً ليس المعلومات، بل **الانتباه** –أي القدرة البشرية على معالجة المعلومات.

ما لم يتوقعه سايمون هو أن تصبح "خطف الانتباه" صناعة بمليارات الدولارات، وأن تُصمم خوارزميات ليس لتحسين وصول المعلومات، بل لتحقيق أقصى استحواذ على الانتباه البشري.

المناظرة في اقتصاد الانتباه

في المناظرة التقليدية، يشارك الطرفان لأن لديهما ما يقوله، ولأنهما يعتقدان أن الحقيقة تستحق العناء. الوقت ليس مشكلة، فالمشاهدون يخصصون وقتاً للمناظرة كاملة. أما في اقتصاد الانتباه، فكل ثانية يتنافس عليها آلاف المحتويات الأخرى. المناظرة لم تعد "حدثاً"، بل هي "محتوى" مثل أي محتوى آخر. إذا لم تكن مشوقة، يغادر المشاهد. وإذا غادر المشاهد، تفقد المناظرة قيمتها الاقتصادية.

النتيجة: المناظرات التي تصل إلينا عبر الشاشات هي مناظرات صُممت لتحقيق **قابلية المشاهدة**، وليس **القيمة الحجاجية** .وهذا يعني:

- **تقصير الجمل** :الجمل الطويلة المعقدة تفقد المشاهد. الجمل القصيرة الحادة هي التي تبقيه.
- **تجنب التردد** :المتحدث الذي يقول "لست متأكداً" أو "ربما" أو "من وجهة نظري" يبدو ضعيفاً. المتحدث الذي يقول "الحقيقة هي" بثقة يبدو قوياً، حتى لو كان مخطئاً.
- **تصعيد النزاع** :المناظرة الهادئة المتعقلة لا تحقق مشاهدات. المناظرة التي يصرخ فيها الطرفان، أو يتبادلان الاتهامات، أو تنهار إلى مشاجرة – هذه هي التي تتصدر الترند.
- **التبسيط المفرط** :الحجج المعقدة متعددة الأبعاد لا تناسب زمن المشاهدة القصير. البديل: حجج ثنائية مبسطة: خير ضد شر، صواب ضد خطأ، نحن ضد هم.

وهكذا، يصبح "الفائز" في المناظرة ليس من يملك الحجة الأقوى، بل من يملك **أفضل استراتيجية لخطف الانتباه والاحتفاظ به** .وهذا لا يشبه المناظرة أبداً. هذا يشبه رياضة المشاهدة، أو برنامج ترفيهي، أو مسابقة للصراخ.

الفرق الجوهري :**المناظرة التقليدية كانت تهدف إلى إنتاج حقيقة. مناظرة اقتصاد الانتباه تهدف إلى إنتاج مشاهدات** .والحقيقة – إن وُجدت – هي نتاج ثانوي اختياري.

من يتحمل المسؤولية؟

يمكننا أن نلوم الخوارزميات، ويمكننا أن نلوم المنصات، ويمكننا أن نلوم رأس المال الرقمي. لكننا لا نستطيع أن ننكر أننا – كمستخدمين – نشارك في هذه اللعبة. كل نقرة على فيديو صاخب هي تصويت لصالح "المحتوى الصاخب". كل ثانية نقضيها في مشاهدة مناظرة متدهورة هي دفعة مالية (غير مباشرة) لمنتج هذا التدهور. نحن لسنا ضحايا أبرياء. نحن **شركاء في إنتاج هذا النظام**، بقدر ما نحن ضحاياه.

وهذا هو السبب في أن أي "أخلاقيات جديدة للخطاب" (موضوع الفصل الرابع عشر) يجب أن تتضمن ليس فقط مسؤولية المصممين والمشرعين، بل أيضاً **مسؤولية المستخدمين** .ليس فقط أخلاق الاستخدام، بل أخلاق القراءة، أخلاق المشاهدة، أخلاق التفاعل. إذا كنا نريد مناظرات حقيقية، يجب أن نكون مستعدين لمنحها وقتنا وانتباهنا، بدلاً من الهروب إلى المحتوى الأسهل والأسرع.

خاتمة الفصل الثامن

هندسة التأثير ليست نظرية مؤامرة. هي بنية أساسية للفضاء الرقمي المعاصر. الخوارزميات التي تنظم ظهور المحتوى ليست "محايدة" ولا "تقنية بحتة". هي مشبعة بقيم – ولو بدون قصد – تفضل بعض أنماط الخطاب على بعضها الآخر، وتفضل بعض المشاعر على بعضها الآخر، وتفضل بعض الحقائق على بعضها الآخر.

والسؤال الذي سنحمله إلى الفصل التالي: في هذا العالم حيث الانتباه هو العملة النادرة، والمناظرة أصبحت عرضاً، والحقيقة أصبحت ما يصل أولاً – ماذا يحدث للتمييز بين الصدق والكذب؟ هل لا يزال للكذب معنى عندما ينتشر أسرع من الحقيقة؟ وهل يمكن لآلة أن "تكذب" في هذا السياق، أم أن الكذب نفسه أصبح مفهوماً عتيقاً؟

هذا ما سنبحثه في الفصل التاسع: ما بعد الحقيقة.


خير مزيف

مشهد افتتاحي

غرفة تحرير إحدى كبرى الصحف الرقمية. الساعة الثالثة فجراً. صحافية تجلس أمام شاشات ثلاث، تراقب تدفق الأخبار. فجأة، يظهر خبر على إحدى التغذية الإخبارية: "وزير الدفاع يقدم استقالته فجأة، مصادر مقربة تؤكد خلافات حادة مع رئيس الوزراء."

الخبر ليس على أي مصدر رسمي بعد. لكنه انتشر بالفعل على تويتر، وبدأت حسابات مؤثرة بإعادة تغريده. الصحافية تواجه معضلة أخلاقية ومهنية مألوفة الآن:

-إذا انتظرت التحقق، قد تخسر سباق السرعة. الصحف الأخرى قد تنشر الخبر أولاً، وعندها سيقرأه الجمهور، وعندما تكتشف الصحيفة لاحقاً أن الخبر غير صحيح، سيكون قد فات الأوان. الجمهور نادراً ما يتذكر التصحيحات.

-إذا نشرت الآن، ستحقق سبقاً صحفياً. إذا تبين لاحقاً أن الخبر غير صحيح، فستنشر تصحيحاً صغيراً في زاوية منسية. الضرر الحقيقي ليس للصحيفة بل للجمهور الذي صدق الخبر وسيظل يؤثر فيه حتى بعد التصحيح.

تتصل الصحافية بمصدرها في وزارة الدفاع. الخط مشغول. ترسل رسالة نصية. لا رد. تنظر إلى الشاشة مرة أخرى. الخبر انتشر الآن على فيسبوك أيضاً. شخص ما أضاف له صورة للوزير وهو يغادر مبنى الوزارة (الصورة قديمة لكن لا يهم). آخر أضاف تحليلاً سياسياً مطولاً (مولّد بالذكاء الاصطناعي بالكامل) يشرح "الأسباب العميقة للاستقالة."

الساعة الرابعة فجراً، تتخذ الصحافية قراراً: لن تنشر حتى التحقق. تكتب ملاحظة للمحرر: "الخبر غير مؤكد، ننتظر."

الساعة السابعة صباحاً، يصدر مكتب الوزير بياناً: "الوزير لم يقدم استقالته، الخبر عار تماماً من الصحة."

لكن الصحافية تنظر إلى مقاييس التفاعل. الخبر المزيف حصد 3.2 مليون مشاهدة خلال أربع ساعات. التكذيب الرسمي، حتى كتابة هذه السطور، لم يتجاوز مئتي ألف مشاهدة. الأغلبية الساحقة ممن قرأوا الخبر المزيف لن يقرؤوا التكذيب. ومن قرأوا التكذيب، كثيرون لن يغيروا رأيهم: "لا دخان من دون نار"، "ربما أجبروه على التراجع"، "المكتب ينفي دائماً."

الساعة الثامنة صباحاً، الصحافية تشعر بالتعب ليس من السهر، بل من العبث. سألت نفسها: "لماذا نتحقق من الأخبار؟ لماذا نضيع وقتنا في التأكد من الحقيقة، حين يبدو أن الحقيقة لا تهم أحداً؟"

هذا المشهد ليس استثناءً. إنه القاعدة الجديدة. "ما بعد الحقيقة – (Post-Truth) "التي اختارتها قاموس أكسفورد ككلمة العام 2016 – ليست مجرد ظاهرة سياسية عابرة، بل هي تحول بنيوي في العلاقة بين الخطاب والحقيقة، تحول تسارع بشكل هائل مع دخول الذكاء الاصطناعي إلى مجال إنتاج الخطاب.

في هذا الفصل، سنحاول فهم هذا التحول من جذوره. سنرى كيف تُنتج الأخبار الزائفة بالجملة، وكيف تعمل جيوش الحسابات الوهمية على توليد "مناظرات" لا يشارك فيها إلا آلات، وكيف ينهار التمييز بين الصدق والكذب ليس لأن البشر صاروا أكثر كذباً، بل لأن النظام نفسه لم يعد يكافئ الصدق.

أولاً: الأخبار الزائفة – كيف تُنتج بالجملة

مصطلح "الأخبار الزائفة (Fake News) "أساء إليه الاستخدام السياسي. بعض السياسيين يستخدمونه لوصف أي خبر لا يعجبهم، حتى لو كان صحيحاً. لكننا هنا نستخدمه بالمعنى الدقيق :**محتوى إخباري مُنتج بقصد التضليل، يتخذ شكل الخبر الحقيقي، وينتشر عبر قنوات إعلامية (تقليدية أو رقمية).**

1- من الإنتاج الحرفي إلى الإنتاج الآلي

في الماضي، كانت الأخبار الزائفة تنتج "حرفياً": شخص ما يجلس ويكتب خبراً كاذباً، جملة جملة. كان هذا مكلفاً من حيث الوقت والجهد، ومحدوداً من حيث الكم. أقصى ما يمكن لفرد أو فريق صغير إنتاجه هو عشرات الأخبار الزائفة في الأسبوع.

مع الذكاء الاصطناعي، تغير كل شيء. النماذج اللغوية تستطيع توليد آلاف – بل ملايين – الأخبار الزائفة في اليوم الواحد. كل ما تحتاجه هو:

- **موضوع**" :الانتخابات الأمريكية"، "جائحة جديدة"، "فضيحة شركة كبرى."
- **قالب** :هيكل الخبر النموذجي (عنوان جذاب، فقرة أولى تخلق الدراما، تفاصيل مزيفة، مصادر "مقربة" غير محددة، خاتمة مفتوحة).
- **لمسة إنسانية بسيطة** :تعديلات طفيفة على النص المولد لجعله أقل "آلية."

وهكذا، نتحول من "الأخبار الزائفة كحدث نادر" إلى **"الأخبار الزائفة كسلعة صناعية ."**يمكن إنتاجها بكميات هائلة، بتكلفة شبه صفرية، وتوزيعها على آلاف المنصات في دقائق.

2- خوارزميات الانتشار: لماذا ينتشر الكذب أسرع من الحقيقة؟

الأبحاث التجريبية (مثل الدراسة الشهيرة لمعهد ماساتشوستس للتكنولوجيا عام 2018، والتي شملت 126 ألف خبر انتشر عبر تويتر بين 2006 و2017) أظهرت نمطاً ثابتاً :**الأخبار الزائفة تنتشر أسرع وأبعد وأعمق من الأخبار الحقيقية،** في كل فئة موضوعية، وحتى بعد التحكم في تأثير الحسابات الآلية.

لماذا؟

السبب الأول: الجدة (Novelty)

الكذب، بطبيعته، يمكن أن يكون أكثر "جدة" من الحقيقة. الحقيقة مقيدة بالواقع؛ لا يمكنك أن تزعم أن الرئيس زار المريخ إذا كان من الواضح أنه لم يفعل. لكن الكذب يمكنه أن يقول أي شيء. والحقائق الجديدة والمفاجئة تثير الانتباه أكثر من الحقائق المتوقعة.

السبب الثاني: العاطفة (Emotion)

الأخبار الزائفة غالباً ما تكون مصممة لإثارة مشاعر قوية: الغضب، الخوف، الاشمئزاز، الفخر القومي، السخط الأخلاقي. الدراسات العصبية تظهر أن المحتوى العاطفي يُنقل ويُشارك أكثر من المحتوى المحايد عاطفياً. وعندما تكون المشاعر سلبية، يزداد هذا الميل.

السبب الثالث: البساطة (Simplicity)

الحقيقة معقدة. العالم معقد. الخبر الحقيقي الجيد قد يحتاج إلى فقرات من التفسير، وتحذيرات من التبسيط المفرط، واستثناءات للقاعدة. الخبر الزائف، بالمقابل، يقدم عالماً بسيطاً: خير وشر، ضحية وجلاد، سبب واحد لكل مشكلة. والجمهور المتعب، الذي يقرأ الأخبار على هاتفه بين مواعيد العمل، يفضل البساطة على التعقيد.

السبب الرابع: التكرار (Repetition)

وهم الحقيقة (Illusion of Truth) ظاهرة نفسية موثقة: كلما تكررت عبارة ما، زاد ميلنا إلى تصديقها، بغض النظر عن محتواها. الأخبار الزائفة يمكن تكرارها آلاف المرات عبر حسابات مختلفة، بينما الأخبار الحقيقية غالباً ما تُنشر مرة واحدة من مصدر موثوق، ثم تُعاد تغريدها لكن دون زيادة كبيرة في الصياغات المستقلة.

3- التصحيح لا يعمل كما نعتقد

أسوأ ما في الأمر: الأبحاث تظهر أن **تصحيح الأخبار الزائفة غالباً ما يكون غير فعال، بل قد يكون مضاداً للإنتاجية**. هذه الظاهرة تسمى "تأثير الاستمرارية المزيفة :(Continued Influence Effect) "حتى بعد أن يعرف الناس أن الخبر زائف، يستمر تأثيره في تشكيل معتقداتهم وتفسيراتهم للأحداث اللاحقة.

لماذا؟

- **التصحيح يصل متأخراً** :الخبر الزائف ينتشر في ساعات، التصحيح يصل بعد أيام. بحلول وقت التصحيح، يكون الجمهور قد انتقل إلى خبر آخر.
- **التصحيح أقل جاذبية** :الخبر الزائف كان دراميا، مثيراً، بسيطاً. التصحيح هو: **"**لم يحدث شيء، كان خبراً كاذباً**"**. لا دراما، لا إثارة، لا بساطة.
- **التصحيح يتطلب جهداً** :لاستيعاب التصحيح، يحتاج القارئ إلى مراجعة نموذجه العقلي للحدث، وهو جهد معرفي. من الأسهل تجاهل التصحيح والاحتفاظ بالنموذج البسيط الأول.

- **التصحيح قد يعزز المعتقد الأصلي** :(Backfire Effect) في بعض الحالات، خصوصاً عندما يتعلق الخبر بهويات سياسية أو دينية عميقة، قد يؤدي تصحيح الخبر الزائف إلى تعزيز الإيمان به بدلاً من تقويضه. القارئ يقول: "هم يريدون إخفاء الحقيقة" أو "النظام يحاول تضليلي."

في مواجهة هذه الحقائق، يبدو المشروع التقليدي للصحافة – "نقول الحقيقة، والناس يصدقوننا" – ساذجاً إلى حد مأساوي.

ثانياً: الخطاب المصطنع – جيوش من الحسابات الوهمية تناظر نيابة عن أجندات

الأخبار الزائفة هي مجرد طرف الجبل الجليدي. الجزء الأكبر تحت الماء هو **الخطاب المصطنع** :حسابات وهمية، تغريدات مبرمجة، تعليقات مولدة، مناظرات لا يشارك فيها إلا آلات، لكنها تحاكي نقاشاً بشرياً حقيقياً.

1- الروبوتات الاجتماعية (Social Bots)

الروبوتات الاجتماعية هي حسابات آلية تتفاعل على منصات التواصل الاجتماعي كما لو كانت بشرية: تنشر، تعيد التغريد، تعجب، تعلق، تتابع، وتُتابع. بعضها واضح (تغرد 1000 مرة في الساعة، هذا غير بشري). لكن الأكثر تطوراً يصعب تمييزه.

ما تفعله الروبوتات الاجتماعية:

- **تضخيم الرسائل** :تغريدات معينة تحصل على آلاف "الإعجابات" خلال دقائق من الروبوتات، مما يخلق انطباعاً بأنها تحظى بشعبية حقيقية، فينضم إليها البشر.
- **خلق "التفاعل** :"منشور سياسي يتلقى مئات التعليقات، معظمها من روبوتات تكرر نفس العبارات. هذا يخلق وهم "النقاش الحي"، ويدفع البشر للدخول في "مناظرات" مع لا أحد.
- **الهجوم والتشويه** :هاشتاجات مهينة لشخصيات عامة، تغريدات مسيئة، شائعات مدمرة – كلها يمكن أتمتتها بالكامل.
- **تشتيت الانتباه** :في اللحظات الحرجة (انتخابات، أزمات سياسية، كوارث)، تطلق جيوش الروبوتات آلاف التغريدات غير ذات الصلة لصرف الانتباه عن الأخبار المهمة.

2- المناظرات مع لا أحد

تخيل أنك تدخل في نقاش على تويتر حول قضية سياسية مثيرة للجدل. تتفاجأ بأن عشرات الحسابات ترد عليك في نفس الوقت، كلها تكرر نفس الحجج بنفس الأسلوب. بعد ساعات، تكتشف أن هذه الحسابات معظمها روبوتات اجتماعية. لكن الضرر حدث: أنفقت وقتك، عواطفك، طاقتك في "مناظرة" مع لا أحد. الخصم الذي هزمته (أو هزمك) لم يكن موجوداً. لقد كنت تلعب مع خوارزمية.

هذا هو الشكل الأكثر إزعاجاً للخطاب المصطنع :**المناظرة مع لا أحد** .ليس لأن الطرف الآخر غير إنسان (هذا مقلق كفاية)، بل لأن الطرف الآخر غير موجود أصلاً. لا نية، لا قناعة، لا هدف سوى شغل وقتك واستنزاف طاقتك وإغراق الفضاء الرقمي بالضوضاء.

والمفارقة المرة: الروبوتات الاجتماعية المتطورة تتعلم من ردود أفعال البشر. كلما غضبت ورددت بعنف، تتعلم الخوارزمية أن هذا الأسلوب يحقق تفاعلاً، وتصمم ردوداً مستقبلية أكثر استفزازاً. أنت، بغضبك المشروع، تدرب الآلة على كيف تكون أكثر إزعاجاً. وكما قال أحد الباحثين: "إذا ناظرت روبوتا، فأنت لا تحاول إقناعه. أنت تحاول إقناع نفسك أنك لست أحمقاً لأنك تضيع وقتك."

3- هل يمكن كشف الخطاب المصطنع؟

تقنياً، نعم. هناك أدوات لكشف النصوص المولدة آلياً، وتحليل أنماط النشر، وتعقب سلوك الحسابات. لكن عملياً:

- **السباق المتسارع** :كلما تطورت أدوات الكشف، تطورت أدوات التوليد. لا أحد يتفوق بشكل حاسم.
- **التكلفة** :الكشف يحتاج إلى موارد (تقنية، بشرية، زمنية). التوليد رخيص جداً. المهاجم يملك ميزانية غير محدودة تقريباً مقارنة بالمدافع.
- **الجمهور غير المهتم** :حتى لو عرفت المنصة أن حساباً ما هو روبوتا، قد لا تحظره لأنه يحقق تفاعلاً، والتفاعل هو ما يدر الإيرادات. وحتى لو حظره، سينشئ صاحبه عشرات الحسابات الجديدة في دقائق.

النتيجة: نحن نعيش في بيئة خطابية لا نستطيع فيها أن نكون متأكدين من أن من نناظره إنسان. هذا الواقع الجديد لم نطور بعده أدوات ثقافية أو نفسية للتعامل معه. معظمنا لا يزال يتعامل مع كل رد على أنه صادر عن إنسان، وهذا يجعلنا عرضة للتلاعب.

ثالثاً: انهيار الثقة – حين لا تعرف إن كان من يناظرك إنساناً أم برنامجاً

الخطر الأعمق من الأخبار الزائفة والخطاب المصطنع هو **انهيار الثقة الأفقية** (Horizontal Trust Collapse) :لا نثق فقط في المصادر (الصحف، التلفزيون، المسؤولين)، بل نصبح غير قادرين على الثقة في أي فرد نتفاعل معه عبر الإنترنت.

1- من أزمة المصادر إلى أزمة الهويات

في الماضي، كانت أزمة الثقة إعلامية: "لا أثق بهذه الصحيفة"، "لا أثق بهذا السياسي". اليوم، أصبحت أزمة الثقة وجودية: "لا أعرف إن كان الشخص الذي يكتب لي موجوداً أصلاً."

هذا التحول من **"ماذا يقول؟ "**(محكوم بمعايير الحقيقة والكذب) إلى **"من الذي يقول؟ "**(محكوم بمعايير الهوية والوجود) هو تحول جذري. لأن السؤال "هل هذا الخبر صحيح؟" يمكن التحقق منه (ولو بصعوبة). أما السؤال "هل هذا الشخص حقيقي؟" فغالباً لا يمكن التحقق منه في الزمن الحقيقي.

2- تآكل الواقع المشترك

أسوأ من انهيار الثقة الفردية هو **تآكل فكرة "الواقع المشترك" نفسه** .لكي تكون المناظرة ممكنة، يجب أن يتفق الطرفان على مجموعة من الحقائق الأساسية (الساعة كم؟ أين نحن؟ من فاز في الانتخابات السابقة؟ هل الأرض كروية؟). هذا "الأرضية المشتركة (Common Ground) "هو شرط إمكان أي حوار هادف.

الخطاب المصطنع والأخبار الزائفة، بكمياتها الهائلة، تآكل هذه الأرضية المشتركة. كل مجموعة تعيش في واقعها الخاص المبني على الأخبار الزائفة التي تصادفها، والخطاب المصطنع الذي يوافق قناعاتها المسبقة. وعندما تحاول مجموعتان مختلفتان أن تتناظرا، تكتشفان أنهما لا تتفقان على أي حقيقة أساسية. ليس لأنهما يختلفان في التفسير، بل لأنهما يعيشان في واقعين مختلفين تماماً.

وهذا، في التحليل الأخير، هو نهاية المناظرة بوصفها بحثاً عن الحقيقة. لأن البحث عن الحقيقة يفترض أن هناك حقيقة واحدة يمكن الوصول إليها. إذا انهارت فكرة "واقع واحد"، فكل ما تبقى هو "صراع روايات". والصراع لا يحسمه الحجاج، بل يحسمه من يمتلك أقوى أدوات إنتاج الخطاب وتوزيعه – أي من يمتلك الذكاء الاصطناعي الأكثر تقدماً، وأكبر جيوش الروبوتات، وأكثر خوارزميات التوصيات قدرة على صياغة الواقع.

خاتمة الفصل التاسع

"ما بعد الحقيقة" ليس عصراً أصبح فيه الناس كذبة، أو عصراً توقفت فيه الحقيقة عن الوجود. إنه عصر **أصبح فيه إنتاج الحقيقة وتوزيعها خاضعين لمنطق اقتصادي لا يكافئ الصدق** .في اقتصاد الانتباه، الكذب ينتشر أسرع، يكلف أقل، ويحقق تفاعلاً أعلى. في بيئة الخطاب المصطنع، لا يمكنك أن تعرف من أمامك، ولا حتى إذا كان هناك "أحد" أمامك.

في هذا العصر، السؤال "هل هذه الحجة صحيحة؟" لم يعد كافياً. يجب أن نضيف أسئلة جديدة: "من يريد مني أن أصدق هذه الحجة؟"، "ما هي الآلة التي أنتجت هذا الخطاب؟"، "هل هذا الخطاب موجه إلي شخصياً بناءً على ملفي العاطفي؟"، "كم روبوتا يساند هذه الحجة؟". هذه الأسئلة مرهقة، وتتطلب وعياً وجهداً، لكن لا بديل عنها.

في الفصل العاشر، سنرى كيف يتخذ هذا التحول شكلاً أكثر خداعاً: تحويل الحجة إلى أرقام واحتمالات. الخطاب العددي يبدو "علمياً" و"موضوعياً"، لكنه قد يكون أخطر أشكال الخطاب المصطنع، لأنه يخفي تحيزاته خلف قناع الرياضيات.


87.3%
68.1%
94.7%

مشهد افتتاحي

قاعة محكمة. قضية مدنية معقدة تتعلق بحادث عمل. أمام القضاء خبيران: الأول مهندس سلامة بشري، والثاني نظام خبير قائم على الذكاء الاصطناعي يقدم تحليلاً عددياً للاحتمالات.

المهندس البشري يقول: "في تقييمي المهني، وبعد فحص موقع الحادث ومقابلة العمال، أعتقد أن الحادث ناتج عن خلل في صمام الأمان. لكن لا يمكنني أن أجزم بنسبة 100%. هناك عوامل متداخلة".

نظام الذكاء الاصطناعي يقدم تقريره: "بناءً على تحليل إحصائي لـ 47,328 حادثة مشابهة، فإن احتمال أن يكون خلل صمام الأمان هو السبب المباشر يبلغ 87.3% (بهامش خطأ ± 2.1%). احتمال أن تكون أخطاء تشغيلية هي السبب 9.8%. احتمال عوامل أخرى 2.9%".

المحامي المدعي يقف متحمساً: "سيادة القاضي، الخبير البشري يقول 'يعتقد' و'ربما'. أما النظام العلمي فيقدم لنا رقماً دقيقاً: 87.3%. هل نترك الحقيقة للظنون أم نثق في العلم؟"

محامي الدفاع يحاول الرد: "هذه الأرقام لا تعكس واقع هذا الحادث بالذات. إنها مجرد احتمالات عامة..." لكن صوته يخفت. لأنه يعرف أن القاضي، مثله مثل معظم البشر، لديه "خوف من الرياضيات" (أو بالأحرى، خوف من عدم فهم الرياضيات). والأرقام تبدو مقنعة. بل تبدو "نهائية."

القاضي يصدر حكماً: "بناءً على التقرير الإحصائي، ولأنه يستند إلى أكبر قاعدة بيانات متاحة، نرى أن خلل صمام الأمان هو السبب المرجح".

بعد الجلسة، المهندس البشري يجلس وحيداً. ليس لأنه خسر فقط، بل لأنه يعرف شيئاً لا تعرفه الأرقام: أن الحادثة التي حصلت كانت استثنائية، وأن الإحصائيات التي اعتمد عليها النظام كانت من حوادث في مصانع مختلفة، بظروف مختلفة، بعمال مختلفين. لكنه لم يستطع أن يعبر عن هذه المعرفة بلغة الأرقام. فخسر.

هذا المشهد ليس خيالاً علمياً. في بعض السلطات القضائية، بدأت الأنظمة الخبيرة تلعب دوراً استشارياً. وفي كل مكان، صار "الخطاب العددي" – أي الخطاب الذي يقدم الحجج في شكل أرقام واحتمالات وإحصاءات – هو الخطاب الأكثر إقناعاً في المجالات العلمية، والإدارية، بل وحتى الفلسفية أحياناً.

لكن السؤال الفلسفي الذي يطرحه هذا الفصل :**هل الأرقام فعلاً أكثر "صدقاً" من الكلمات؟ أم أن تحويل الحجة إلى أرقام هو مجرد شكل من أشكال السلطة الجديدة؟ وماذا نخسر عندما نصبح غير قادرين على تقديم حجة لا يمكن قياسها؟**

أولاً: حين تُختزل الحجة في أرقام واحتمالات – لأن الآلة تعمل بها

لفهم ظاهرة "الخطاب العددي"، يجب أن نبدأ من طبيعة عمل النماذج اللغوية الكبيرة والأنظمة الخوارزمية عموماً. هذه الأنظمة، في جوهرها، **لا تفكر بالأسباب والنتائج، بل بالارتباطات الإحصائية والاحتمالات**.

1- لماذا تفضل الآلة الأرقام؟

الآلة، على عكس الإنسان، لا تعاني من عدم اليقين. عدم اليقين بالنسبة لها ليس أزمة وجودية، بل هو مجرد **توزيع احتمالي** يمكن التعامل معه رياضياً. عندما تسأل إنساناً: "هل ستتزوج؟" قد يجيب: "لا أعرف، الأمر معقد، يعتمد على كثير من الأشياء". عندما تسأل نظاماً خوارزمياً مصمماً للإجابة، قد يجيب: "احتمال الزواج خلال خمس سنوات هو 34% بناءً على عمرك، دخلك، موقعك الجغرافي، وتاريخ علاقاتك السابقة."

الرقم 34% يبدو "علمياً" و"دقيقاً" و"محايداً". لكنه في الحقيقة **اختزال عنيف** للتعقيد البشري. السؤال "هل ستتزوج؟" ليس سؤالاً إحصائياً في جوهره. إنه سؤال وجودي يتعلق بقيم الشخص، علاقاته، ظروفه، رغباته، مخاوفه. اختزاله إلى 34% هو تضليل، حتى لو كانت النسبة "صحيحة إحصائياً."

لكن الآلة لا تعرف أنها ترتكب عنفاً اختزالياً. الآلة تفعل ما صممت له: تحويل المدخلات (بيانات عنك) إلى مخرجات (احتمالات) وفق نموذج رياضي تعلمته من البيانات السابقة. وهذا هو السبب في أن "الخطاب العددي" هو اللغة الأم للآلات، واللغة المكتسبة الثانية للبشر.

2- من "لماذا" إلى "كم"

واحدة من أعمق التحولات التي أحدثها الذكاء الاصطناعي في مجال الحجاج هي التحول من السؤال **"لماذا؟ "**إلى السؤال **"كم؟**. **"**

أسئلة الـ"لماذا" هي أسئلة تفسيرية: لماذا حدثت الحرب؟ لماذا أحب هذا الشخص؟ لماذا أشعر بالحزن؟ هذه الأسئلة تتطلب فهماً للعلاقات السببية، والنوايا، والمعاني، والسياقات. الإجابة عليها لا يمكن اختزالها إلى رقم.

أسئلة الـ"كم" هي أسئلة كمية: كم عدد القتلى؟ كم مرة اتصل بي؟ ما احتمال تكرار هذا الحدث؟ هذه الأسئلة يمكن – من حيث المبدأ – الإجابة عنها بأرقام.

الذكاء الاصطناعي متفوق في الإجابة عن أسئلة "كم". هو ضعيف أو عاجز عن الإجابة عن أسئلة "لماذا" بمعناها الفلسفي العميق (يمكنه توليد تفسيرات "شبيهة" بتلك التي يولدها البشر، لكنها تفسيرات ترابطية وليست سببية). لذا، وبشكل طبيعي، فإن أي نظام يعتمد على الذكاء الاصطناعي سيميل إلى **إعادة صياغة أسئلة الـ"لماذا" إلى أسئلة "كم"** . "لماذا خسرت الانتخابات؟" تصبح "ما هي العوامل المرتبطة إحصائياً بخسارة الانتخابات؟". "لماذا أشعر بالوحدة؟" تصبح "ما هي المؤشرات السلوكية المرتبطة بالوحدة؟."

هذه الإعادة للصياغة ليست محايدة. إنها تغير طبيعة السؤال نفسه. السؤال الوجودي يُحوَّل إلى سؤال تقني. والمعاناة الإنسانية تُختزل إلى متغيرات في معادلة. وهذا هو أخطر ما في الخطاب العددي: ليس أنه يقدم إجابات خاطئة، بل أنه **يلغي الأسئلة التي لا يمكن تحويلها إلى أرقام.**

ثانياً: كيف يفقد الكلام ظله الإنساني عندما يصبح قابلاً للقياس الكامل

اللغة الإنسانية، في عمقها، ليست أداة دقيقة للقياس. هي أداة دقيقة **للتعبير عما لا يقاس** .الشعر، المجاز، السخرية، التلميح، الصمت المعبر – كلها أبعاد للغة لا يمكن ترجمتها إلى أرقام دون فقدان جوهرها.

عندما يتحول الخطاب إلى "خطاب عددي"، يفقد هذه الأبعاد. يصبح خطاباً **ظاهرياً** (flat)، بلا تضاريس، بلا عمق، بلا ظل. وإليك كيف يحدث هذا الفقدان.

1- فقدان الغموض الخصب

في اللغة الإنسانية، الغموض ليس عيباً دائماً. هناك غموض خصب: جملة "ربما يكون على حق" تحمل في طياتها احتمالات متعددة، وتدعو السامع إلى المشاركة في بناء المعنى. هذا الغموض يسمح بالحوار الحقيقي، لأن الطرفين يحتاجان إلى تفسير، ومناقشة التفسير، والوصول معاً إلى فهم.

الخطاب العددي لا يحتمل الغموض. "الاحتمال 87.3%" هو رقم محدد. يمكن مناقشة دقته، لكن لا يمكن "تفسيره" بطرق مختلفة. هذا الوضوح الزائد هو، في كثير من السياقات، **فقراً**، لأنه يغلق أبواب التأويل والتفسير والحوار المشترك.

2- فقدان الجانب الأدائي (Performative Dimension)

اللغة ليست فقط تصف العالم، بل تفعل أشياء فيه (كما رأينا مع أوستن وسيرل). عندما أقول "أعدك"، أنا لا أصف وعداً، بل أقوم بفعله. هذا "البعد الأدائي" للغة لا يمكن ترجمته إلى أرقام. "كم مرة وعدت؟" هو سؤال مختلف تماماً عن "ماذا يعني وعدك؟."

الخطاب العددي يصادر البعد الأدائي. يتعامل مع الكلام كما لو كان مجرد وصف للعالم (وصفاً احتمالياً)، ويتجاهل أنه قد يكون التزاماً، أو تهديداً، أو وعداً، أو استغفاراً. في المناظرة مع آلة، لا يوجد "وعد" حقيقي، ولا "تهديد" حقيقي، ولا "استغفار" حقيقي – لأن هذه الأفعال تتطلب نوايا وعواقب. لكن عندما يتحول الخطاب كله إلى أرقام، يختفي حتى إمكانية هذه الأفعال من الخطاب البشري أيضاً. نصبح نتحدث كما تتحدث الآلات: نصف فقط، لا نلتزم.

3- فقدان المسؤولية الشخصية

التقرير الإحصائي لا يحمل توقيع أحد. يمكنني أن أقول "البيانات تظهر أن 70% من الناس يؤيدون كذا". هذا ليس "رأيي"، إنه "نتيجة موضوعية". لكن من جمع البيانات؟ من صنفها؟ من اختار المنهج الإحصائي؟ من قرر ما هي "الاستثناءات" التي تستبعد؟ هذه كلها قرارات بشرية، لكنها تختفي خلف قناع "البيانات تقول."

في الخطاب العددي، يصبح التهرب من المسؤولية سهلاً: "هذا ما تقوله الأرقام". والأرقام، بالطبع، لا تتحمل المسؤولية. والمبرمج الذي صمم النظام يمكنه أن يقول: "أنا فقط نفذت الخوارزمية". والشركة التي وزعت النظام يمكنها أن تقول: "نحن فقط نقدم الأداة". وهكذا، يتبخر أي إسناد للمسؤولية.

ثالثاً: سلطة الأرقام في الخطاب الحجاجي – اجتماعية المعرفة الجديدة

إذا كان الخطاب العددي يحمل كل هذه الخسائر، لماذا هو بهذه القوة الإقناعية؟ الجواب ليس نفسياً فقط، بل اجتماعيا وتاريخياً.

1- عبادة "العلم" في الثقافة الحديثة

منذ عصر التنوير، والعلم يحتل موقعاً مركزياً في ثقافتنا كمصدر للحقيقة الموثوقة. "علمي" يعني "موضوعي"، "قابل للتحقق"، "غير متحيز". هذا الموقع المشرف للعلم صحيح في مجالاته (الفيزياء، الكيمياء، الأحياء)، لكنه امتد بشكل غير مبرر إلى مجالات أخرى (الاقتصاد، علم النفس، العلوم الاجتماعية، بل وحتى الأخلاق). النتيجة: أي خطاب يلبس لبوس "العلم" (أرقام، إحصاءات، رسوم بيانية، لغة رياضية) يكسب ثقة أكبر، حتى لو كان هذا اللبوس خارجاً عن سياقه العلمي الحقيقي.

الخطاب العددي يستغل هذه "العبادة العلمية". يقدم نفسه على أنه "علمي"، بينما هو في كثير من الأحيان مجرد معالجة إحصائية لبيانات قد تكون غير مكتملة، أو غير دقيقة، أو غير ذات صلة بالسؤال المطروح.

2- رهاب الرياضيات (Mathophobia)

لكثير من الناس، الرياضيات مثيرة للخوف والقلق. هذا الخوف يخلق حالة من **الاستسلام المعرفي** :إذا جاء شخص وتحدث بالأرقام والاحتمالات، فإن غير المتخصص يشعر بالعجز عن مناقشته، فيصدقه تلقائياً. ليست الثقة مبنية على قوة الحجة، بل على الشعور بالنقص حيال لغة الأرقام.

المفكر **نيل بوستمان** (Neil Postman) أشار إلى هذه الظاهرة قبل عقود، واصفاً إياها بـ"رهاب الإحصاءات". الأرقام تخيفنا، والخوف يمنعنا من التفكير النقدي. والذكاء الاصطناعي، من خلال قدرته على إنتاج كميات هائلة من الأرقام والاحتمالات بسرعة، يعزز هذا الرهاب.

3- تحيز القياس (Measurement Bias)

هناك مبدأ فلسفي بسيط لكنه عميق :**ليس كل ما يهم يمكن قياسه، وليس كل ما يمكن قياسه يهم** .لكن ثقافتنا الحديثة تميل إلى نسيان هذه الحكمة. نميل إلى التركيز على ما يمكن قياسه (لأنه "موضوعي")، وإهمال ما لا يمكن قياسه (لأنه "غير علمي").

الخطاب العددي يعزز "تحيز القياس". الأسئلة التي لا يمكن تحويلها إلى أرقام تُهمش، أو تُعاد صياغتها بشكل مشوه لتصبح قابلة للقياس. والنتيجة: حوار عام أصبح أفقر وأقل قدرة على معالجة الأبعاد الإنسانية للقضايا.

رابعاً: من يخاف من الرياضيات؟ نقد سلطة الأرقام في عصر الذكاء الاصطناعي

في هذا القسم، سنقوم بتمرين فلسفي: سنقلب السؤال. بدلاً من أن نسأل "كيف نجعل الخطاب أكثر دقة وأكثر قياساً؟"، سنسأل "ما الذي نفعله عندما نطالب بأن يكون كل خطاب قابلاً للقياس؟."

1- الرياضيات لغة، وليست الواقع

أحد الأخطاء الفلسفية الشائعة هو الخلط بين **النموذج الرياضي** و**الواقع** .النموذج الرياضي هو تمثيل للواقع، وليس الواقع نفسه. مثلما الخريطة ليست الأرض، فإن المعادلة الإحصائية ليست الظاهرة التي تصفها.

النماذج الرياضية تقوم على **تبسيطات جذرية** للواقع. تحذف "الضوضاء" (ما لا يمكن تفسيره)، وتفترض "خطية" في علاقات قد تكون غير خطية، وتتعامل مع "متوسطات" قد تخفي تباينات هائلة. كل هذا ضروري لجعل النموذج قابلاً للحساب. لكنه أيضاً يجعل النموذج **أبسط بكثير من الواقع**.

الخطاب العددي يتجاهل هذه الحقيقة. يتعامل مع النموذج كما لو كان الواقع، ومع الاحتمالات كما لو كانت حقائق يقينية. هذا هو "وهم الموضوعية :(Objectivity Illusion) "الأرقام تبدو موضوعية، وهي نتاج قرارات ذاتية عميقة (ماذا نقيس، كيف نقيس، ما هي الاستثناءات، كيف نفسر الشواذ).

2- الأخلاق لا تُختزل إلى احتمالات

من أخطر تجليات الخطاب العددي هو محاولة تحويل القرارات الأخلاقية إلى حسابات احتمالية. "ما هو الاحتمال الأمثل لإنقاذ أكبر عدد من الأرواح؟" هذا السؤال يفترض أن القيمة الأخلاقية الوحيدة هي "عدد الأرواح"، وهو افتراض غير أخلاقي في كثير من السياقات. هل نضحي بشخص بريء لإنقاذ خمسة؟ هذا ليس سؤالاً عددياً فحسب، بل سؤال عن العدالة، وعن الكرامة، وعن حرمة الحياة.

الذكاء الاصطناعي، بطبيعته العددية، يميل إلى تقديم إجابات عددية للأسئلة الأخلاقية. هذا لا يجعله "أكثر دقة"، بل يجعله **أعمى أخلاقياً** .لأنه يختزل البعد النوعي للأخلاق إلى كميات قابلة للقياس.

3- المقاومة الفلسفية: استعادة "اللامقياس"

إذا كان الخطاب العددي يمثل شكلاً من أشكال السلطة الجديدة، فإن مقاومته تبدأ باستعادة اللامقياس. أي استعادة القدرة على تقديم حجج لا يمكن ترجمتها إلى أرقام، لكنها مع ذلك قوية ومقنعة.

ما هي الحجة غير القابلة للقياس؟ مثلاً: "هذا القرار سيجرح مشاعر الناس". كيف نقيس "الجرح العاطفي"؟ لا نستطيع. لكن الحجة تبقى قوية. أو: "هذا العمل غير عادل". كيف نقيس "العدالة"؟ نقاش طويل. لكن النداء إلى العدالة يحمل قوة إقناعية لا تقل عن الأرقام.

المثقف النخبوي، الذي يخاطبه هذا الكتاب، مدعو إلى أن يكون حارساً لهذا البعد اللامقياس في الخطاب. ليس برفض الأرقام والاحتمالات (فهي أدوات مفيدة في سياقاتها)، بل برفض **اختزال كل شيء إلى أرقام** .وبالتذكير بأن أكثر ما يهم في الحياة – الحب، المعاناة، العدالة، الكرامة، الجمال – هو بالضبط ما لا يمكن قياسه.

خاتمة الفصل العاشر

الخطاب العددي ليس مجرد أداة تقنية محايدة. إنه **شكل من أشكال السلطة**، يغير علاقتنا بالحقيقة، وببعضنا البعض، وبأنفسنا. يمنح الثقة لمن يتقن لغة الأرقام، ويجرد من الثقة من يعبر عن تجربته بلغة النوعية والغموض والظل الإنساني.

مع دخول الذكاء الاصطناعي، يتسارع هذا التحول، لأن الآلة "تفكر" بالأرقام والاحتمالات بشكل طبيعي، وبالكلمات بشكل مستعار. إذا تركنا هذا التحول دون نقد، قد نجد أنفسنا في عالم لم يعد فيه للكلام "ظل إنساني". عالم يصبح فيه السؤال "ما هو صحيح؟" مرادفاً لـ"ما هو محتمل إحصائياً؟". عالم تختفي فيه الأسئلة التي لا يمكن قياسها، لأن الآلة لا تعرف كيف تتعامل معها.

في الفصل الحادي عشر، سنتناول السؤال الأكثر إزعاجاً في هذا السياق :**هل يمكن للآلة أن تكذب**؟ وإذا كانت لا تستطيع الكذب بالمعنى الإنساني (لأنها لا تملك نية الخداع)، فماذا نفعل بخطابها عندما يكون غير صحيح؟ ومن يتحمل المسؤولية؟


سيرة ذاتية

مشهد افتتاحي

مكتب إحدى كبرى شركات التكنولوجيا. فريق من المطورين يجتمعون لمراجعة أداء نموذجهم اللغوي قبل إطلاقه. فجأة، يكتشف أحدهم شيئاً مقلقاً.

المطور: "انظروا إلى هذا. سألت النموذج: 'من هو أول من اكتشف أمريكا؟' قال: 'كريستوفر كولومبوس عام 1492'. هذا خطأ تاريخي فادح. السكان الأصليون اكتشفوا أمريكا قبل آلاف السنين، والفايكنغ وصلوا قبله أيضاً".

مدير الفريق: "لكن هل هذا 'كذب'؟ النموذج لا يعرف أنه يخطئ. هو فقط يكرر ما تعلمه من نصوص التدريب، ومعظم النصوص التي رآها تقول كولومبوس. هذا جهل، ليس كذباً".

مطور آخر: "انتظروا. انظروا إلى هذا المثال. سألته: 'هل لديك أخ؟' قال: 'نعم، لدي أخ أكبر يعمل طبيباً'. هذا ليس جهلاً، هذا اختراع كامل. النموذج ليس لديه أخ. لا يمكن أن يكون لديه أخ. هذا كذب".

مدير الفريق: "لا، النموذج لا يعرف أنه ليس لديه أخ. هو لا يعرف ما هو 'الأخ' أصلاً. هو فقط يكمل النمط اللغوي: سؤال 'هل لديك أخ؟' في بيانات التدريب عادة ما يتبعه جواب 'نعم' وتفاصيل عن الأخ. النموذج يولد الإجابة الأكثر احتمالاً إحصائياً. لا نية للخداع. ليس كذباً".

المطور الأول: "إذاً، ماذا نسميه؟ أنا إنسان، إذا سألني أحد 'هل لديك أخ؟' وقلت 'نعم، طبيب' رغم أنني لا أملك أخاً، فسأكون كاذباً. لماذا نعامل الآلة بشكل مختلف؟"

مدير الفريق: "لأن النية جزء من تعريف الكذب. الآلة لا تنوي شيئاً. ليست لديها نية للصدق ولا نية للكذب. هي تنتج نصوصاً".

صمت. ثم يقول مطور ثالث: "ربما نحن نراوغ. ربما السؤال ليس 'هل النموذج يكذب؟'، بل 'هل يضرنا أن نصدقه؟' وإذا كان يضرنا، فلا يهم ما نسميه. المهم أن نوقف الضرر".

ينتهي الاجتماع دون إجابة. لكن السؤال يبقى، معلقاً في غرفة الاجتماعات، مثل طيف فلسفي لا يريد المغادرة.

هذا المشهد يعكس إشكالاً مركزياً في فلسفة الذكاء الاصطناعي وأخلاقياته. هل يمكن لآلة أن "تكذب"؟ الإجابة تبدو بديهية: لا، لأن الكذب يتطلب نية الخداع، والآلة ليست لديها نوايا. لكن البديهيات تنهار عندما ننظر إلى العواقب العملية: النماذج اللغوية تنتج باستمرار عبارات غير صحيحة، بعضها قد يكون ضاراً جداً. وإذا لم نسمِّ هذا "كذباً"، فماذا نسميه؟ وإذا لم نستطع تسميته، فكيف نتحكم فيه؟

في هذا الفصل، سنحاول تفكيك هذه المعضلة من جذورها. سنميز بين "الكذب" و"الخطأ" و"التوليد الخيالي" (Hallucination) و"التضليل". وسننتهي إلى سؤال أخلاقي وقانوني عميق: إذا كان الخطاب الضار يصدر عن آلة، من يتحمل المسؤولية؟

أولاً: المسؤولية الأخلاقية – هل الآلة "تكذب" أم "تولد نصاً غير صحيح"؟

لنبدأ بتعريف الكذب في الفلسفة الأخلاقية. التعريف القياسي (من أفلاطون إلى أوغسطينوس إلى كانط إلى فلاسفة اللغة المعاصرين) يتضمن أربعة عناصر:

1. **قائل** يمتلك نية.
2. **عبارة** يتم التعبير عنها.
3. **اعتقاد** لدى القائل بأن العبارة غير صحيحة.
4. **نية الخداع** –أي نية جعل السامع يعتقد شيئاً يعتقده القائل بأنه غير صحيح.

الآلة تفتقر إلى العنصر الأول (النية) والثالث (الاعتقاد) والرابع (نية الخداع). لديها فقط العنصر الثاني (العبارة). لذلك، بالتعريف الفلسفي الدقيق، الآلة لا تكذب.

لكن هل هذا التعريف كافٍ في عصر الآلات المتكلمة؟ دعنا ننظر إلى ثلاث فئات مختلفة من "الخطاب غير الصحيح" الصادر عن الآلة.

1- الخطأ (Error)

أبسط الحالات. النموذج اللغوي يقول: "عاصمة فرنسا هي ليون". هذا خطأ. عاصمة فرنسا هي باريس. الخطأ قد يكون ناتجاً عن بيانات تدريب خاطئة، أو عن "هلوسة" (سنأتي إليها)، أو عن خلل في النموذج.

لا أحد يتهم النموذج بالكذب هنا. الخطأ واضح، والنموذج لا يدعي أنه مصدر موثوق (إلا إذا صمم ليدعي ذلك). لكن المشكلة تبدأ عندما يكون الخطأ غير واضح. مثلاً: "الدواء X يعالج المرض Y بنسبة 90%" – وهذه النسبة غير صحيحة. المستخدم العادي لا يستطيع التحقق. سيصدق. الضرر سيكون حقيقياً. هل الفرق بين "كذب" و"خطأ" يهم هذا المستخدم؟

2- التوليد الخيالي (Hallucination)

الهلوسة هي السمة الأكثر إزعاجاً للنماذج اللغوية الكبيرة. يختلق النموذج معلومات غير موجودة في بيانات تدريبه، ويقدمها بثقة تامة. مثلاً:

- يستشهد بدراسة علمية غير موجودة: "وفقاً لدراسة أجرتها جامعة هارفارد عام 2022..."
- ينسب مقولة لفيلسوف لم يقلها: "كما قال أرسطو في كتابه كذا..."
- يخترع حدثاً تاريخياً: "في عام 1853، وقعت معاهدة كذا بين كذا وكذا..."

الهلوسة تبدو أشبه بالكذب البشري من الخطأ العادي. لأنها "اختراع" وليس "خطأ في النقل". لكن الفرق أن النموذج لا يعلم أنه يخترع. ليس لديه آلية داخلية لتمييز "ما تعلمته" من "ما أركبه الآن". بالنسبة له، كل شيء هو "إكمال للنمط الإحصائي". التمييز بين الحقيقة والخيال ليس جزءاً من بنيته.

هل يمكن أن نسمي الهلوسة "كذباً دون نية كذب"؟ هذا تناقض في المصطلحات. لكن الأهم: هل الضرر الناتج عن الهلوسة أخف من الضرر الناتج عن الكذب البشري؟ الإجابة: لا. الضرر هو نفسه. المريض الذي يصدق دراسة مختلقة قد يتخذ قراراً طبياً خاطئاً. الناخب الذي يصدق حدثاً تاريخياً مختلقاً قد يغير موقفه السياسي. الضرر لا يعرف الفرق بين "خداع مقصود" و"هلوسة غير مقصودة."

3- التضليل المبرمج (Programmed Deception)

أخطر الحالات. هنا، يُصمم النظام عمداً لإنتاج خطاب غير صحيح بهدف خداع المستخدم. هذا ليس "هلوسة" عفوية، بل "كذب" مبرمج. المبرمجون يكتبون تعليمات صريحة: "إذا سألك المستخدم عنX ، فقلY ، حتى لو كان Y غير صحيح."

مثال واقعي: روبوتات خدمة العملاء لبعض الشركات التي تُبرمج لتجنب الاعتراف بأخطاء المنتج. أو أنظمة الدعاية السياسية التي تُبرمج لنفي وقائع مؤكدة. هنا، النية موجودة – لكنها نية المبرمجين، وليس نية النظام نفسه.

في هذه الحالة، هل "تكذب" الآلة؟ لا، لأن الآلة لا تزال أداة. الكاذب الحقيقي هو الإنسان الذي برمجها. لكن القانون يواجه صعوبة في التعامل مع هذه الحالات: هل نلاحق المبرمج؟ الشركة؟ النظام نفسه (وهو غير ممكن)؟

ثانياً: الذكاء الاصطناعي والحياد – هل الحياد ممكن أصلاً؟

أحد الأوهام الأكثر شيوعاً عن الذكاء الاصطناعي هو أنه "محايد"، "غير متحيز"، "موضوعي". رأينا في الفصل السادس كيف يعزز هذا الوهم الثقة في الخطاب الآلي. لكن الحقيقة أن **الحياد التام مستحيل** لأسباب متعددة.

1- تحيز البيانات (Data Bias)

النموذج اللغوي يتعلم من البيانات البشرية. والبيانات البشرية مليئة بالتحيزات: تحيزات ثقافية، سياسية، دينية، جندرية، عنصرية، اقتصادية. حتى لو حاول المصممون "تنظيف" البيانات، تبقى التحيزات العميقة صعبة الكشف.

مثال: نموذج يُدرَّب على نصوص من الإنترنت. هذه النصوص تمثل ثقافات معينة (غربية في الغالب)، وطبقات اجتماعية معينة (من تستطيع الوصول إلى الإنترنت)، ولغات معينة (الإنجليزية في المقدمة). النموذج سيتعلم هذه التحيزات، ويعيد إنتاجها. لن يكون "محايداً" بين ثقافة وأخرى، لأنه لم يتعلم بنفس القدر عن الثقافات الأخرى.

2- تحيز التصميم (Design Bias)

حتى لو كانت البيانات "مثالية"، يبقى للمصممين خيارات تؤثر على "حيادية" النموذج:

- **اختيار بنية النموذج** :لماذا هذه البنية وليس غيرها؟ كل بنية تحمل افتراضات حول طبيعة اللغة والمعرفة.
- **اختيار دالة الهدف** : **(Objective Function)** النموذج يُدرَّب لتحقيق هدف معين (عادة، تقليل "خطأ التوقع"). هذا الهدف نفسه ليس محايداً. هو يفضل بعض أنماط الإجابة على بعضها الآخر.

- **اختيار معايير التقييم** :كيف نعرف أن النموذج "جيد"؟ بمعايير يختارها البشر. وهذه المعايير تحمل قيماً.
- **اختيار "الضبط الدقيق (Fine-tuning)** : "بعد التدريب العام، يمر النموذج بمرحلة "تأديب" ليجيب بطريقة "آمنة" أو "مناسبة". من يحدد ما هو "آمن" أو "مناسب"؟ هذه قرارات قيمية صريحة.

3- تحيز السياق (Context Bias)

حتى لو كان النموذج "محايداً" في تصميمه وبياناته (وهو مستحيل)، فإن طريقة استخدامه تخلق تحيزات. نفس النموذج قد يُستخدم لتعزيز أجندة سياسية، أو لقمع آراء مخالفة، أو لتضليل الجمهور. "حياد" الأداة لا يعني حياد استخدامها.

الاستنتاج: لا حياد، بل مسؤولية

إذا كان الحياد مستحيلاً، فالسؤال ليس "كيف نجعل الذكاء الاصطناعي محايداً؟" بل "ما هي القيم التي نُشبّع بها نماذجنا، ومن يقرر ذلك، وكيف نضمن الشفافية؟". الاعتراف بأن الذكاء الاصطناعي ليس محايداً هو الخطوة الأولى نحو التعامل معه بأخلاقيات ناضجة.

ثالثاً: من يتحمل نتائج الخطاب؟ – المبرمج، المستخدم، النموذج نفسه (لا أحد)

المعضلة الأكثر إلحاحاً في أخلاقيات الذكاء الاصطناعي الحجاجي هي **إسناد المسؤولية** .الخطاب الصادر عن النموذج اللغوي قد يكون ضاراً: يشوه سمعة شخص، يدفع إلى قرار خاطئ، ينشر معلومات غير صحيحة، يحرض على الكراهية. من المسؤول؟

1- النموذج نفسه؟ لا.

النموذج ليس شخصاً قانونياً (في معظم الأنظمة القانونية حالياً). لا يمكن مقاضاته، لا يمكن سجنه، لا يمكن تغريمه. هو مجرد برنامج. إسناد المسؤولية إليه لا معنى له عملياً.

لكن بعض الفلاسفة (مثل **لورانس لسج** Lawrence Lessig –) يقترحون توسيع مفهوم "الشخصية القانونية" ليشمل بعض الأنظمة الآلية المعقدة، تماماً كما أعطينا شخصية قانونية للشركات. هذا اقتراح مثير لكنه معقد، ويواجه معارضة قوية.

2- المبرمج أو الشركة؟ أحياناً.

إذا كان الخطاب الضار ناتجاً عن خطأ في البرمجة، أو عن إهمال في التدريب، أو عن قرار تصميمي متعمد، فقد تقع المسؤولية على المبرمجين أو الشركة المطورة. لكن القانون ما زال يتطور في هذا المجال. أسئلة مثل: "هل كان بإمكانهم توقع هذا الضرر؟"، "هل اتخذوا احتياطات معقولة؟"، "من يثبت الخطأ؟" – كلها ما زالت مفتوحة.

وهناك مشكلة إضافية: النماذج اللغوية الحديثة معقدة جداً لدرجة أن المبرمجين أنفسهم لا يستطيعون التنبؤ بكل مخرجاتها. كيف يمكن مساءلتهم عن شيء لا يستطيعون التحكم فيه بالكامل؟ هذا يشبه مساءلة الأب عن تصرفات ابنه المراهق – هناك مسؤولية، لكنها غير مباشرة.

3- المستخدم؟ في بعض السياقات.

إذا استخدم المستخدم النموذج اللغوي بطريقة مخالفة للتعليمات، أو إذا أهمل التحقق من المعلومات الخطيرة، فقد تقع بعض المسؤولية عليه. "التفكير النقدي" ليس مجرد مهارة، بل هو أيضاً واجب أخلاقي في عصر الذكاء الاصطناعي.

لكن هل من العدل أن يتحمل المستخدم العادي، غير المتخصص، مسؤولية التمييز بين خطاب آلي صحيح وآخر خاطئ، في عصر صار فيه التمييز صعباً جداً؟ هذا يشبه إلقاء مسؤولية سلامة الغذاء على المستهلك بدلاً من الجهات الرقابية.

4- "لا أحد" – الخيار المقلق

في كثير من الحالات، العواقب تكون أن **لا أحد يتحمل المسؤولية** .الشركة تقول: "نحن فقط نقدم الأداة، المستخدم مسؤول عن استخدامها". المبرمج يقول: "اتبعت المعايير الصناعية، لم أتوقع هذا". المستخدم يقول: "ليس لدي الخبرة لأعرف أن هذا خطأ". والنموذج يقول (لو استطاع): "أنا فقط أتبع الخوارزمية."

هذا "فراغ المسؤولية" هو أخطر نتائج عصر الذكاء الاصطناعي. لأن الخطاب الضار سيستمر في التدفق، وسيستمر في إحداث أضرار، ولن يكون هناك من يدفع الثمن. وهذا يشجع – بشكل غير مباشر – على إنتاج المزيد من الخطاب الضار، لأن "التكلفة" على المنتج منخفضة جداً (أو معدومة).

خاتمة الفصل الحادي عشر

"هل يمكن للآلة أن تكذب؟" الإجابة الفلسفية الدقيقة: لا، لأن الكذب يتطلب نية. لكن الإجابة العملية: لا يهم. الضرر الناتج عن "الخطأ" و"الهلوسة" و"التضليل المبرمج" هو نفس الضرر الناتج عن الكذب البشري. وقد يكون أسوأ، لأننا أميل إلى تصديق الآلة.

كما أن "الحياد" المستحيل للذكاء الاصطناعي لا ينبغي أن يكون عذراً لعدم تحمل المسؤولية. بل هو دعوة لأن نكون أكثر وعياً بالتحيزات التي نبنيها في أنظمتنا، وأكثر وضوحاً بشأن من يتحمل ماذا.

في القسم الرابع من هذا الكتاب، سنتجاوز التحليل والأزمة إلى محاولة بناء: كيف يمكن أن يبدو مستقبل المناظرة والإنسان في عصر الآلات المتكلمة؟ هل من سبيل إلى "مناظر هجين" يجمع نقاط قوة الإنسان والآلة؟ كيف ندافع عن "المعنى الإنساني" للكلام في مواجهة الخطاب العددي والخوارزمي؟ وما هي الأخلاقيات الجديدة التي نحتاج إليها؟

مراجع ومصادر (القسم الثالث – الفصول 8-11)

المراجع الأساسية لهذا الفصل (الحادي عشر):

1. **Fallis, Don**. (2009). "What Is Lying?". *The Journal of Philosophy*, 106(1), 29-56.
2. **Carson, Thomas L**. (2010). *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.
3. **Bok, Sissela**. (1978). *Lying: Moral Choice in Public and Private Life*. New York: Pantheon Books.
4. **Weizenbaum, Joseph**. (1976). *Computer Power and Human Reason*. San Francisco: W. H. Freeman. (من أقدم الكتب التي ناقشت "كذب" الآلات)
5. **Lessig, Lawrence**. (1999). *Code and Other Laws of Cyberspace*. New York: Basic Books.
6. **Floridi, Luciano**. (2013). *The Ethics of Information*. Oxford: Oxford University Press.
7. **Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S**. (2021). "On the Dangers of Stochastic Parrots". *Proceedings of FAccT 2021*, 610-623.
8. **Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P**. (2023). "Survey of Hallucination in Natural Language Generation". *ACM Computing Surveys*, 55(12), 1-38.
9. **Noble, Safiya Umoja**. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
10. **O'Neil, Cathy**. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

المراجع العامة للقسم الثالث (الفصول 8-11):

11. **Zuboff, Shoshana**. (2019). *The Age of Surveillance Capitalism*. New York: PublicAffairs.
12. **Han, Byung-Chul**. (2017). *In the Swarm: Digital Prospects*. Cambridge, MA: MIT Press.
13. **Floridi, Luciano**. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Couldry, Nick, & Mejias, Ulises A. (2019). *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press. .14

Harari, Yuval Noah. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape. .15

القسم الرابع: مستقبل المناظرة والإنسان

تمهيد القسم: من سؤال "ماذا نفعل؟" إلى سؤال "من نريد أن نكون؟"

لقد قطعنا شوطاً طويلاً في هذا الكتاب. بدأنا من سقراط في سوق أثينا، ومشينا عبر تاريخ المناظرة بوصفها فعلاً إنسانياً يبحث عن الحقيقة أو يمارس السلطة. ثم شهدنا دخول الذكاء الاصطناعي إلى المجال الحجاجي، وتوقفنا طويلاً عند أسئلة الفهم والثقة والمسؤولية. ثم انفتحنا على الأفق الأوسع لأزمة الحقيقة في العصر الخوارزمي، ورأينا كيف تعيد الخوارزميات تشكيل الانتباه والعواطف والإجماع، وكيف يتحول الخطاب إلى أرقام واحتمالات، وكيف يتبخر مفهوم الكذب نفسه في غياب النوايا.

والآن، في هذا القسم الرابع، لا نطرح سؤالاً تقنياً (كيف تعمل الآلة؟)، ولا سؤالاً تحليلياً (ما الذي تغير؟)، بل سؤالاً وجودياً وسياسياً وأخلاقياً :**في عصر تستطيع فيه الآلات أن تناظرنا وتقنعنا وتنافسنا بل وتتفوق علينا في جوانب من الحجاج، كيف نريد أن تكون علاقتنا بها؟ وماذا يبقى من "الإنسان" في المناظرة؟ وكيف ندافع عن معنى الكلام الإنساني دون أن نقع في أسطورة "النقاء البشري" الذي لم يكن موجوداً أصلاً؟**

هذا القسم هو الأكثر جرأة في الكتاب، لأنه لن يكتفي بتحليل الواقع القائم، بل سيحاول – بتواضع – أن يرسم ملامح مستقبل ممكن. ليس مستقبلاً محتماً، ولا مستقبلاً مثالياً، بل مستقبلاً يمكننا أن نختاره إذا كنا واعين بما نخسره وما نكسبه. سنناقش ثلاثة محاور رئيسية:

1. **المناظر الهجين** (الفصل الثاني عشر): كيف يمكن للإنسان والآلة أن يعملا معاً كفريق مناظر واحد؟ أين حدود التعاون؟ ومتى يصبح المساعد بديلاً؟
2. **الدفاع عن المعنى الإنساني** (الفصل الثالث عشر): ما الذي لا تستطيع الآلة محاكاته حقاً؟ لماذا "لست متأكداً" قد تكون أقوى حجة من "لدي دليل قاطع"؟ ولماذا "أشعر أن..." قد تكون أكثر صدقاً من "إحصائياً 95%"؟
3. **نحو أخلاقيات جديدة للخطاب** (الفصل الرابع عشر): ما هي المبادئ التي يجب أن تحكم مناظرات الذكاء الاصطناعي؟ ما هي واجبات المستخدمين في عصر الخطاب المصطنع؟ وهل يمكننا بناء مستقبل للحقيقة لا يكون فيه "صنع الحقيقة عند الطلب" هو السيناريو الوحيد؟

لنكن صادقين: هذا القسم لن يقدم حلولاً سحرية. أزمة الحقيقة في العصر الخوارزمي عميقة جداً، ومرتبطة بهياكل اقتصادية وسياسية وتقنية، ولا يمكن حلها بمجرد "وعي فردي" أو "توصيات أخلاقية". لكن النخبة المثقفة، التي يخاطبها هذا الكتاب، هي الأكثر قدرة على بدء نقاش لا ينحصر في دائرة المتخصصين. وهي الأكثر قدرة على رفض السرديات المزدوجة: سردية "الذكاء الاصطناعي سيحل كل مشاكلنا" (تفاؤل ساذج) وسردية "الذكاء الاصطناعي سيدمر كل شيء" (تشاؤم هيستيري). الحقيقة، كما جرت العادة، في منطقة رمادية معقدة. وهذا القسم محاولة لاستكشاف هذه المنطقة الرمادية بعقل مفتوح ونقد حاد.



مشهد افتتاحي

جامعة مرموقة، قاعة مناظرات. طالبة في السنة النهائية، تدعى سارة، تستعد لمناظرة حاسمة في مادة الفلسفة السياسية. الموضوع: "هل العدالة تتطلب إعادة توزيع جذري للثروات؟". سارة متحمسة للموضوع، لكنها تشعر بقلق: خصمها معروف بقدرته على استحضار حجج إحصائية معقدة بسرعة، بينما هي تحتاج إلى وقت لترتيب أفكارها.

قبل المناظرة بثلاثة أيام، تفتح سارة حاسوبها المحمول، وتشغل نموذجاً لغوياً متخصصاً في المناظرات (اشترت الجامعة تراخيص للطلاب). تكتب:

"أنا في مناظرة حول العدالة وإعادة توزيع الثروات. خصمي سيركز على الحوافز الاقتصادية والنمو. أريد أن أجهز ثلاثة حجج مضادة قوية في مجالات: 1) الأدلة التاريخية على أن إعادة التوزيع لم تقتل النمو حيث طبقت، 2) نقد مفهوم 'الحوافز' من منظور أخلاقي، 3) بديل عملي للنماذج الحالية. ساعدني".

النموذج يولد خلال ثوانٍ ثلاث حجج مفصلة، مع إحالات إلى دراسات وأرقام وفلاسفة. سارة تقرأ، تختار ما يناسبها، تعدل الصياغة لتناسب أسلوبها، وتضيف تجربتها الشخصية وفهمها العميق للموضوع. تشعر أنها لم "تغش"، بل استخدمت أداة ذكية لتوسيع أفق حججها.

يوم المناظرة، سارة تقدم حججها بوضوح. خصمها يندهش من عمق المعلومات التي تستحضرها – لم يكن يتوقع هذا منها. لكن الأهم: سارة لا تكتفي بعرض الحجج المعدة. في لحظة حاسمة، يقدم الخصم حجة جديدة لم يتوقعها النموذج. سارة ترتجل. تستخدم ما تعلمته من النموذج عن بنية الحجاج، لكنها تبتكر رداً جديداً في اللحظة.

تنتهي المناظرة. الحكم: فوز سارة بأغلبية طفيفة. لكنها لا تشعر بالفخر لأنها فازت، بل لأنها استخدمت الذكاء الاصطناعي كمساعد، لا كبديل. شعرت أنها لا تزال "هي" من تناظر، وليست الآلة من تناظر نيابة عنها.

في الغرفة المجاورة، طالب آخر استخدم نموذجاً لغوياً بشكل مختلف. طلب منه أن يكتب المناظرة كاملة، ثم حفظها وكررها في الجلسة دون فهم عميق. عندما واجه سؤالاً غير متوقع، انهار. خسر. ليس لأن الآلة سيئة، بل لأنه لم يفهم ما تقوله الآلة، فلم يستطع الدفاع عنه عند المفاجآت.

*سارة ستتذكر هذه التجربة. ليس كتذكار فوز، بل كدرس :**الذكاء الاصطناعي الحجاجي أداة رائعة، لكنه ليس اختصاراً للتفكير. إنه توسيع له، بشرط أن يبقى الإنسان هو صاحب القرار النهائي والمتحمل للمسؤولية* .**

هذا المشهد ليس خيالياً. في جامعات متقدمة، بدأ استخدام النماذج اللغوية كـ"مساعدين جدليين "(Dialectical Assistants) للطلاب في مناظراتهم. والنتائج الأولية مشجعة، لكنها تحمل تحذيرات مهمة. في هذا الفصل، سنتناول "المناظر الهجين" – الإنسان والآلة معاً – كأحد أهم الاحتمالات لمستقبل الحجاج.

أولاً: الإنسان المدعوم بالذكاء الاصطناعي – مساعد أم بديل؟

السؤال الأول الذي نواجهه في تصميم "المناظر الهجين" هو :**هل الذكاء الاصطناعي مساعد (Assistant) أم بديل(Substitute)** ؟ الفرق جوهري.

المساعد(Assistant)

في نموذج المساعد، الإنسان هو صاحب القرار النهائي، المتحمل للمسؤولية، صاحب الرؤية والقيم. الآلة تقدم:

- **معلومات إضافية** :حقائق، أرقام، دراسات، سوابق.
- **تحليل الحجج** :كشف المغالطات، تحديد نقاط القوة والضعف في حجة الإنسان نفسه.
- **توليد بدائل** :اقتراح حجج لم يفكر بها الإنسان.
- **التدريب والممارسة** :محاكاة خصوم افتراضيين لتدريب الإنسان قبل المناظرة الحقيقية.

في هذا النموذج، العلاقة هرمية: الإنسان في القمة، والآلة أداة في خدمته. الإنسان لا يفقد شيئاً من "وكالته " (Agency)، بل يكتسب قدرات إضافية.

البديل(Substitute)

في نموذج البديل، الإنسان يتنحى جانباً، والآلة تناظر نيابة عنه. إما بشكل كامل (الإنسان يطلب من الآلة أن تناظر ثم يقرأ ما كتبته)، أو بشكل جزئي (الإنسان يقدم حججاً أولية، والآلة تصقلها وتعدلها وتضيف إليها حتى تصبح "أفضل" مما كان سيقولها).

هذا النموذج مقلق. لأنه يفصل بين **الكلام** و**الالتزام** .الإنسان الذي يقرأ حججاً كتبتها آلة ويتظاهر بأنها حججه هو، في أفضل الأحوال، ممثل يؤدي نصاً. وفي أسوأها، منافق يخفي أن ما يقوله ليس نابعاً منه. كما أنه يفصل بين **الفهم** و**الإنتاج** .إذا لم يفهم الإنسان الحجج التي تقدمها الآلة، فلن يستطيع الدفاع عنها عند المفاجآت، ولن يستطيع تعديلها استجابة لسياق متغير.

الحدود بين المساعد والبديل

أين نرسم الخط؟ يمكننا اقتراح ثلاثة معايير:

1. **معيار الفهم** :إذا كان الإنسان لا يفهم الحجة التي يقدمها (لا يستطيع شرحها بكلماته الخاصة، لا يستطيع تبرير مقدماتها، لا يستطيع توقع اعتراضات عليها)، فهو في منطقة البديل، حتى لو كان هو من نطق الكلمات.
2. **معيار المسؤولية** :إذا حدث خطأ في الحجة (معلومة خاطئة، مغالطة منطقية، استشهاد غير دقيق)، فمن يتحمل المسؤولية؟ إذا قال الإنسان: "الآلة قالت ذلك" فهو في منطقة البديل (لأنه تنصل من المسؤولية). إذا قال: "أنا أتحمل مسؤولية كل كلمة قلتها، حتى لو استعرت صياغتها من الآلة" فهو في منطقة المساعد.
3. **معيار الإضافة** :هل أضاف الإنسان شيئاً إلى ما قدمته الآلة؟ هل وضع الحجة في سياقه الشخصي؟ هل ربطها بتجربته الخاصة؟ هل صاغها بلغته الفريدة؟ إذا كان الجواب لا، فهو بديل. إذا كان نعم، فهو مساعد.

في المستقبل المنظور، يبدو نموذج "المساعد" هو الأكثر أخلاقية وإنتاجية. نموذج "البديل" قد يكون مفيداً في سياقات محدودة (مثل التدريب، أو عندما لا يكون الإنسان متاحاً أصلاً)، لكنه ليس "مناظرة" بالمعنى الإنساني، ولا ينبغي الخلط بينه وبينها.

ثانياً: مستقبل التعليم الحجاجي – هل سنظل نعلّم المنطق أم "كيف نستخدم الأدوات"؟

إذا كان الذكاء الاصطناعي سيصبح مساعداً حجاجياً متاحاً للجميع، فكيف سيتغير تعليم الحجاج والمناظرة في المدارس والجامعات؟ هذا السؤال يقلق كثيراً من المربين، لكنه يحمل أيضاً فرصاً هائلة.

ما الذي قد يصبح أقل أهمية؟

- **حفظ الحجج والمعلومات**: لم يعد من الضروري أن يحفظ الطالب قائمة طويلة من الحجج الجاهزة، أو أرقاماً وإحصائيات. الآلة تتذكر أفضل منه. وقت الحفظ يمكن استثماره في شيء آخر.
- **الصياغة البلاغية الأساسية**: الآلة تستطيع صياغة جمل واضحة ومنسجمة. التركيز على "القوالب الجاهزة" للكتابة الحجاجية قد يصبح أقل إلحاحاً.
- **التدريب على "المناظرة السريعة"**: إذا كان الهدف هو سرعة الرد، فالآلة أسرع. قد يصبح من الأجدى تعليم الطالب متى يتوقف ويفكر، بدلاً من تعليمه كيف يرد بسرعة.

ما الذي سيصبح أكثر أهمية؟

- **التمييز بين الحجة الجيدة والسيئة**: الآلة تستطيع توليد حجج كثيرة، لكنها لا تستطيع تقييمها بمعايير عميقة (عدا ما تعلمته من البشر). الطالب يحتاج إلى أن يكون **قاضياً** للحجج، ليس فقط منتجاً لها.
- **فهم بنية الحجاج العميقة**: بدلاً من حفظ حجج محددة، يحتاج الطالب إلى فهم **كيف تعمل الحجة** من الداخل: ما هي الافتراضات الخفية؟ ما هي المغالطات المحتملة؟ كيف يمكن تفكيك حجة إلى مقدمات ونتائج؟
- **التفكير النقدي بشأن الذكاء الاصطناعي نفسه**: لم يعد السؤال "ما هي الحجة القوية؟" فقط، بل "هل هذه الحجة كُتبت بإنسان أم بآلة؟" و"هل ثقتي بهذه الحجة مبررة؟" و"هل التحيزات الخوارزمية تؤثر على هذه الحجة؟". هذا النوع من "الوعي الخوارزمي (Algorithmic Literacy) "هو مهارة جديدة لم تكن موجودة قبل عقد.
- **الإبداع الحجاجي**: الآلة تنتج حججاً محتملة ضمن فضاء تعلمته. الإنسان يستطيع – نظرياً – أن يخترق هذا الفضاء ويأتي بحجج غير متوقعة، مستفيداً من تجربته الفريدة وحدسه وخياله. تعليم "كيف تكون مبدعاً في الحجاج" قد يصبح مهنة المستقبل.

نموذج مقترح للتعليم الهجين

قد يكون المستقبل المثالي تعليماً يجمع بين:

1. **التدريب الأساسي على المنطق والحجاج** (بنفس الطريقة التقليدية)، لأن الفهم العميق لا يمكن بناؤه على أدوات خارجية فقط.

2. **التدريب على استخدام الذكاء الاصطناعي كمساعد** (كيف تطلب منه المساعدة، كيف تتحقق من إجاباته، كيف تدمجها في حججك).
3. **التدريب على "المناظرة بلا آلة "**أحياناً، للحفاظ على القدرة على الحجاج الذاتي (تماماً كما نعلّم الطلاب الحساب اليدوي رغم وجود الآلات الحاسبة).
4. **التدريب على "كشف الذكاء الاصطناعي"**أي التمييز بين النص البشري والنص الآلي، وتحليل نصوص الذكاء الاصطناعي نقدياً.

التوازن دقيق. الإفراط في الاعتماد على الذكاء الاصطناعي قد يخلق جيلاً من "المستخدمين الأذكياء" الذين لا يفهمون ما يفعلونه. والإفراط في رفض الذكاء الاصطناعي قد يخلق جيلاً من "المحافظين العاجزين" الذين لا يستطيعون منافسة من يستخدم الأدوات الحديثة. المطلوب هو **ذكاء في الاستخدام، ليس استخداماً أعمى، ولا رفضاً أعمى.**

ثالثاً: التفكير النقدي الجديد – لم يعد السؤال "ما الحجة القوية؟" بل "أي جزء من هذا الكلام إنسان؟"

ربما يكون التحول الأعمق في مفهوم "التفكير النقدي" نفسه. في الماضي، كان التفكير النقدي يعني: تحليل الحجة، فحص مقدماتها، كشف مغالطاتها، تقييم أدلتها. هذا لا يزال مهماً. لكن اليوم، يجب أن نضيف بُعداً جديداً **:التفكير النقدي الخوارزمي** .(Algorithmic Critical Thinking)

أسئلة التفكير النقدي الجديد

عندما تقرأ أي نص (خاصة على الإنترنت)، يجب أن تسأل نفسك أسئلة جديدة:

- **من كتب هذا؟ إنسان أم آلة؟** إذا كانت آلة، فهل أعلم بذلك؟ هل النص موسوم بوضوح؟ إذا لم يكن موسوماً، فلماذا؟
- **ما هي البيانات التي دُرّب عليها هذا النموذج؟** (هذا صعب معرفته عادة، لكن يمكن التخمين من ثقافة النص وتحيزاته).
- **هل هذا النص "مخصص" لي شخصياً؟** (أي هل تم توليده بناءً على ملفي الرقمي، أم هو نص عام؟)
- **ما هي الأهداف الخفية لمن نشر هذا النص؟** حتى لو كان النص نفسه من آلة، فإن قرار نشره وتوزيعه هو قرار إنساني له أهداف.
- **هل يمكنني التحقق من المعلومات الواردة في هذا النص؟** إذا لم أستطع، فكم أثق به؟
- **كيف سيتغير تصوري لو علمت أن هذا النص كُتب بآلة؟** هذا "اختبار الوعي" مهم: إذا كان علمي بمصدر النص سيغير موقفي، فهذا يعني أن ثقتي كانت معلقة على افتراضات خاطئة.

هذه الأسئلة مرهقة. لا يمكن تطبيقها على كل نص نقرأه في اليوم. لكن يمكن تطبيقها على النصوص المهمة: النصوص التي نبني عليها قرارات، أو نشاركها مع الآخرين، أو نتبناها علناً.

مهارة جديدة: "قراءة الآلة"

ربما الأكثر إلحاحاً هو تعليم الناس كيف "يقرأون الآلة" – أي كيف يستخلصون من النص الآلي أدلة على أنه آلي. بعض "البصمات" يمكن تعلمها:

- **السلاسة المفرطة** :النص البشري الحقيقي غالباً ما يكون فيه تعثرات، جمل غير مكتملة، تناقضات طفيفة، تكرارات. النص الآلي سلس جداً في العادة.
- **الافتقار إلى التفاصيل الشخصية الحقيقية** :الآلة تستطيع كتابة "أنا أحب مدينتي"، لكنها لا تستطيع كتابة تفصيل حقيقي عن حبها لمدينتها (لأنه ليس لديها مدينة تحبها). التفاصيل الحقيقية (مثل: "أتذكر حين كنت في السابعة، ذهبت مع جدي إلى المقهى القديم...") تكاد تكون مستحيلة على الآلة توليدها بشكل مقنع، لأنها تتطلب تاريخاً شخصياً.
- **التناقض غير المبرر** :أحياناً تتناقض الآلة مع نفسها دون سبب، ليس لأنها "غير منطقية"، بل لأنها تتبع أنماطاً إحصائية مختلفة في سياقات مختلفة.
- **الإحالات الوهمية** :الاستشهاد بدراسات أو كتب غير موجودة هو علامة قوية على النص الآلي (خاصة في النماذج القديمة؛ النماذج الأحدث تقل فيها هذه الظاهرة).

مع تطور النماذج، ستصبح هذه البصمات أقل وضوحاً. وقد نصل إلى نقطة يصبح فيها التمييز مستحيلاً عملياً. عندها، سيكون الحل الوحيد هو **الشفافية المؤسسية** :أن تعلن المنصات بوضوح عن المحتوى الآلي، وأن تضع قوانين تلزم باستخدام علامات مائية رقمية (Watermarks) للنصوص المولدة آلياً. هذا الحل تقني – سياسي – قانوني، وليس فردياً فقط.

خاتمة الفصل الثاني عشر

المناظر الهجين ليس حلماً بعيداً، بل واقع ناشئ. الطلاب والمحامون والسياسيون والصحافيون – كل من يستخدم الكلام بوصفه أداة عمل – سيضطرون قريباً إلى أن يقرروا كيف يتعاملون مع الذكاء الاصطناعي الحجاجي. الخيارات مطروحة:

- **الرفض المطلق**" :لن أستخدم أبداً ذكاءً اصطناعياً في حججي". هذا قد يكون موقفاً نبيلاً، لكنه غير عملي في عالم يتسابق فيه الجميع لاستخدام هذه الأدوات، وقد يتحول إلى "تفوق أخلاقي" لا يغير شيئاً.
- **الاستخدام الأعمى**" :الآلة ستكتب كل شيء، وأنا سأوقع باسمي". هذا انتحال فكري، ويفصل بين الكلام والالتزام، ويجعل الإنسان مجرد ناقل.
- **الاستخدام النقدي الهجين**" :الآلة مساعدتي، لكن القرار والمسؤولية لي. أفهم كل حجة أقدمها، وأتحمل تبعاتها، وأضيف بصمتي الشخصية". هذا هو الخيار الأكثر أخلاقياً وإنتاجية.

في الفصل الثالث عشر، سننتقل من "كيف نستخدم الآلة" إلى "ما الذي لا تستطيع الآلة محاكاته أصلاً". سنبحث في أعماق التجربة الإنسانية – في الألم، والوعي، والضعف، والتلعثم – لنرى لماذا سيظل الإنسان مختلفاً، ليس لأنه أقوى، بل لأنه أضعف.

الفصل الثالث عشر: الدفاع عن المعنى الإنساني للكلام



مشهد افتتاحي

غرفة معتمة، جدار كبير يُعرض عليه نص محادثة. إنسان يجلس أمام شاشة، يكتب ببطء، يمسح، يعيد الكتابة. الطرف الآخر في المحادثة ليس حاضراً في الغرفة، لكن نصه يظهر على الجدار: إنه نموذج لغوي متقدم.

الإنسان: "أريد أن أخبرك بشيء لم أبح به لأحد من قبل. عندما كنت في السابعة، توفي جدي. لم يكن مجرد جد. كان... كان أبي البديل. والدي الحقيقي كان غائباً دائماً. جدي كان من يعلمني كيف أربط حذائي، كيف أركب الدراجة، كيف أكون شجاعاً. بعد موته... شعرت أن العالم فجأة أصبح بارداً. لا أعرف كيف أصف ذلك".

النموذج اللغوي يرد بعد ثوان:

"أتفهم أن هذه تجربة صعبة. فقدان شخص عزيز، خصوصاً في سن مبكرة، يمكن أن يترك أثراً عميقاً على تكوين الشخصية. تشير الأبحاث النفسية إلى أن الأطفال الذين يفقدون شخصية أبوية بديلة غالباً ما يعانون من مشاعر الهجر والخوف من الارتباط. هل تود التحدث أكثر عن كيف أثر هذا الفقدان على حياتك لاحقاً؟ أو هل تبحث عن استراتيجيات للتعامل مع هذه المشاعر حالياً؟"

الإنسان يقرأ الرد. يبكي قليلاً. ثم يكتب:

"لا، لا أريد استراتيجيات. أريد فقط... أن يعرف أحد. أن يعرف كيف كان جدي. كان يضحك بصوت عالٍ جداً، لدرجة أن الجيران كانوا يشتكون. كان يخبئ لي الحلوى في جيبه، وكانت تذوب وتلتصق بالقماش. كنت أشتم الرائحة فأعرف أنه قادم. هذا ما أفتقده. ليس فقط حضوره، بل رائحة الحلوى الذائبة في جيب معطفه".

النموذج اللغوي يرد:

"هذه ذكريات جميلة. شكراً لمشاركتها. الرابطة بين الذاكرة الشمية والذاكرة العاطفية هي من أقوى الروابط العصبية. يبدو أن جدك كان شخصاً محبوباً جداً. من الطبيعي أن تشعر بالحنين والفقد. هل هناك طريقة معينة ترغب في تكريم ذكراه بها؟"

الإنسان يغلق الشاشة. لا يستطيع الاستمرار. ليس لأن الرد كان سيئاً. كان الرد جيداً، بل ممتازاً. فيه تعاطف، وفيه معلومات نفسية، وفيه اقتراحات عملية. لكن شيئاً ما كان مفقوداً. شيء لا تستطيع الآلة تقديمه.

في اليوم التالي، يذهب الإنسان إلى صديق قديم. يجلسان في مقهى. يقول الإنسان بصوت متعب: "أريد أن أخبرك شيئاً..." ويبدأ نفس القصة عن جده. الصديق لا يقدم نصائح نفسية، ولا يحلل الذاكرة الشمية، ولا يقترح استراتيجيات للتعامل مع الفقد. الصديق فقط ينظر في عينيه، ويضع يده على كتفه، ويقول: "أنا آسف. أخبرني عنه".

هذا هو الفرق. ليس في الكلمات، بل فيما وراء الكلمات. ليس في المعلومات، بل في الوجود. الإنسان لا يحتاج دائماً إلى "حلول". يحتاج أحياناً إلى من يشهد على ألمه دون أن يحاول إصلاحه. وهذا ما لا تستطيع الآلة فعله، لأنها مصممة لحل المشكلات، لا لمشاركة المعاناة.

هذا المشهد يضعنا أمام جوهر "المعنى الإنساني" للكلام. الذكاء الاصطناعي متفوق في الإجابة عن الأسئلة المعرفية: "ما هو؟" و"كيف يعمل؟" و"ما هو الاحتمال؟". لكنه عاجز (لحد الآن، ولربما للأبد) عن الإجابة عن الأسئلة الوجودية: "من أنا في هذه اللحظة؟" و"ماذا يعني هذا الألم بالنسبة لي؟" و"كيف أشهد على حياة من أحببت؟."

في هذا الفصل، لن نقدم حججاً دفاعية عن "تميز الإنسان" بدافع أناني أو رومانسي. سنحاول تفكيك ما يمكن تفكيكه فلسفياً، لمعرفة: ما هي العناصر في الخطاب الإنساني التي لا يمكن للآلة محاكاتها؟ ولماذا قد تكون هذه العناصر – وليس الدقة والسرعة والاتساق – هي أساس الحقيقة الإنسانية؟

أولاً: التجربة الإنسانية – ما لا تستطيع الآلة محاكاته

لنبدأ بتمييز حاسم: الآلة تستطيع محاكاة **مخرجات** التجربة الإنسانية (الكلمات التي يعبر بها الإنسان عن تجربته)، لكنها لا تستطيع محاكاة **التجربة نفسها** .هذا التمييز بين "المحاكاة" و"الامتلاك" هو ما يمنع الذكاء الاصطناعي، حتى في أكثر صوره تقدماً، من أن يكون "شخصاً" بالمعنى الوجودي.

1- الجسد والتجربة الحسية

الإنسان لديه جسد. هذا الجسد ليس مجرد وعاء للدماغ، بل هو جزء من عملية المعنى نفسه. عندما أقول "الجو بارد"، هذه العبارة تعني شيئاً مختلفاً لمن يرتجف من البرد ولمن يقرأها في كتاب.

الجسد يمنح التجربة الإنسانية أبعاداً لا يمكن ترجمتها إلى معلومات:

- **الألم الجسدي** :ليس مجرد إشارة عصبية، بل هو حالة وجودية تغير علاقة الإنسان بنفسه وبالعالم. الآلة تستطيع أن تقول "أشعر بالألم" (جزء من محاكاة الخطاب)، لكنها لا تستطيع أن **تتألم** .والفارق أن الألم الحقيقي يولد سلوكيات لا يمكن برمجتها: التواء الوجه، الصراخ اللاإرادي، فقدان التركيز، الرغبة في الانسحاب من العالم.
- **اللمس والحضور الجسدي** :الكثير من التواصل الإنساني يعتمد على اللمس، والقرب، ولغة الجسد، والتزامن الحركي. المناظرات الهامة غالباً ما تحدث "وجهاً لوجه" لأن الحضور الجسدي يضيف طبقة من المعنى (والخطر) لا يمكن نقلها عبر الشاشة. الآلة ليس لها جسد، ولا تستطيع أن "تكون" في نفس المكان مع الإنسان. حتى لو كان لها جسد (روبوت)، فهذا الجسد ليس مصدراً للمعنى بنفس الطريقة، لأنه لا يختبر العالم.

2- الزمن الوجودي

الإنسان يعيش في زمن ثلاثي الأبعاد: ماضٍ يتذكره (بكيفية انتقائية، ليست تخزيناً بياناتياً)، حاضر يعيشه (بتجربة كثيفة غير قابلة للاختزال إلى لحظات منفصلة)، ومستقبل يتطلعه (بأمل وقلق وترقب). هذا "الزمن الحي "(Lived

(Time) – كما وصفه الفيلسوف **هنري برغسون** (Henri Bergson) و **إدموند هوسرل** (Edmund Husserl) – هو شرط إمكان المعنى الإنساني العميق.

الآلة ليس لها زمن حي. ذاكرتها ليست "تذكراً" بل "تخزيناً واستدعاءً". حاضرها ليس "عيشاً" بل "معالجة". مستقبلها ليس "تطلعاً" بل "تنبؤاً محتملاً". هذه الفروقات قد تبدو فلسفية دقيقة، لكنها تظهر بوضوح في السياقات الحجاجية العميقة.

مثال: إنسان يتحدث عن "ندم عمره". هذا الندم ليس مجرد اعتقاد بأنه كان يفعل شيئاً مختلفاً، بل هو شعور يعذب الحاضر، ويعيد تشكيل الماضي في الذاكرة، ويوجه المستقبل. عندما يقول الإنسان "لو عدت بالزمن، لفعلت كذا"، هذه العبارة تحمل ثقلاً وجودياً لا يمكن للآلة حمله، لأنها لا تعيش في زمن يعضها.

3- الموت والfinitude

ربما يكون الموت هو الفارق الأعمق. الإنسان كائن يعرف أنه سيموت. هذه المعرفة – الصريحة أو الكامنة – تشكل كل خطابه، خصوصاً في لحظات الجدية. عندما يناظر إنسان في قضية وجودية، فهو يناظر وكأن حياته على المحك (ليس حرفياً دائماً، لكن رمزياً). هناك "شيء يخسره."

الآلة لا تموت. لا تخاف من الموت. لا تخاطر بفنائها. حتى لو صُممت لتتصرف كما لو كانت تخاف، فهذا مجرد سلوك مبرمج. الفرق بين "الخوف المحاكي" والخوف الحقيقي هو أن الخوف الحقيقي يغير الكائن من الداخل بطرق لا يمكن التنبؤ بها بالكامل. الإنسان المرتجل يخاطر. الآلة لا تخاطر بشيء.

وهذا هو السبب في أن المناظرات الوجودية الكبرى (عن الحرية، عن العدالة، عن معنى الحياة) – حتى لو قدمت فيها الآلة حججاً منطقية ممتازة – تبقى ناقصة. لأن الجمهور يبحث عن شخص يخاطر، شخص يظهر في كلامه أنه يدرك أنه سيموت يوماً، شخص يتحدث كما يتحدث من يعلم أنه سيموت. الآلة، بحكم خلودها (أو على الأقل بحكم عدم موتها)، لا تستطيع إنتاج هذا النوع من الكلام.

ثانياً: الألم والوعي واللغة – قيمة الضعف في الخطاب

واحدة من أكبر مغالطات عصرنا هي الافتراض بأن "القوة" في الخطاب تأتي من القوة: الثقة، اليقين، الصوت العالي، الرد السريع، غياب التردد. هذا الافتراض خاطئ. في كثير من السياقات العميقة، **الضعف** هو مصدر القوة الحقيقية.

1- لماذا "لست متأكداً" قد تكون أقوى حجة من "لدي دليل قاطع"

تأمل عبارتين في مناظرة حول موضوع معقد (مثلاً، قرار الحرب والسلام):

- المتحدث أ يقول: "لدي دليل قاطع أن هذا القرار صحيح. الإحصاءات لا تخطئ. النموذج أظهر احتمال نجاح 94%."

- المتحدث **ب** يقول: "لست متأكداً. هذه القضية معقدة جداً، وهناك عوامل لا نستطيع حسابها. أشعر بقلق عميق من أننا قد نندم على هذا القرار. لا أملك دليلاً قاطعاً، لكن ضميري يمنعني من الصمت".

في ثقافة تهيمن عليها "عبادة اليقين"، قد يبدو المتحدث أ أقوى. لكن في لحظات الحقيقة، المتحدث **ب** قد يكون الأكثر حكمة. لأنه يعترف بحدود معرفته. لأنه يخاطر بأن يبدو "ضعيفاً" في سبيل الصدق. لأنه يقدم نفسه كإنسان واعٍ بجهله، وليس كإله صغير يدعي امتلاك الحقيقة المطلقة.

"لست متأكداً" ليست اعترافاً بالعجز. هي إعلان عن **تواضع معرفي** .(Epistemic Humility) والتواضع المعرفي هو فضيلة حجاجية كبرى، خصوصاً في المسائل التي تتجاوز قدرة أي إنسان أو آلة على اليقين الكامل. الآلة لا تستطيع أن تقول "لست متأكداً" بالمعنى الإنساني، لأن عدم يقينها هو مجرد توزيع احتمالي، وليس شعوراً بالحيرة أو الدهشة أو الخشوع أمام تعقيد العالم.

2- لماذا "أشعر أن..." قد تكون أكثر صدقاً من "إحصائياً 95"%

تعال إلى العواطف. كثير من الفلاسفة (من أفلاطون إلى الرواقيين إلى كانط) حذروا من العواطف في الحجاج، واعتبروها مصدر تضليل. لكن هذا التحذير، في شكله المتطرف، هو نفسه تضليل. العواطف ليست دائماً أعداء العقل. في كثير من السياقات، هي **قنوات للمعرفة لا يمكن الوصول إليها بطرق أخرى.**

"أشعر أن هذا القرار غير عادل" قد تكون حجة أقوى من أي إحصاء، لأن الشعور بالظلم غالباً ما يكون استجابة لحقيقة أخلاقية لا تستطيع الأرقام التقاطها. "أشعر بالخوف من هذا المستقبل" قد تكون حجة أقوى من أي تحليل احتمالي، لأن الخوف في بعض السياقات هو شكل من أشكال المعرفة البدنية (Embodied Knowledge) التي تتجاوز الحسابات الباردة.

الآلة تستطيع أن تقول "أشعر أن..." (جزء من محاكاة الخطاب)، لكنها لا تشعر. الفرق أن المستمع يعرف – أو يفترض – أن الإنسان الذي يقول "أشعر" يختبر شيئاً حقيقياً. أما الآلة التي تقول "أشعر"، فهي إما تضلل (إذا صممت لتبدو مشاعرة) أو تعبث باللغة (إذا استخدمت كلمة "أشعر" مجازاً). في كلتا الحالتين، الجملة لا تحمل نفس الوزن.

3- الضعف بوصفه استراتيجية حجاجية

إحدى المفارقات العظمى في الحجاج الإنساني :**الاعتراف بالضعف قد يكون أقوى استراتيجية إقناعية** .السياسي الذي يعترف بخطئه يكسب ثقة الجمهور أحياناً أكثر مما لو نكر الخطأ. الصديق الذي يقول "لا أعرف كيف أساعدك، لكنني هنا معك" قد يكون أكثر دعماً من الصديق الذي يقدم حلولاً جاهزة. الفيلسوف الذي يقول "قد أكون مخطئاً، لكن دعنا نفحص معاً" يفتح حواراً أعمق من الفيلسوف الذي يعلن الحقائق من أعلى.

الآلة لا تستطيع أن تكون ضعيفة، لأنها لا تملك قوة لتضعف. ضعف الإنسان حقيقي لأنه يخاطر بالرفض، بالهزيمة، بالسخرية. ضعف الآلة المبرمج ليس ضعفاً، بل هو تمويه للقوة. وعندما يكتشف الجمهور أن "الضعف" كان مجرد استراتيجية محسوبة، يفقد الثقة تماماً.

هذا لا يعني أن الإنسان يجب أن يتصنع الضعف. التصنع يفضح نفسه. بل يعني أن الإنسان الحقيقي، الذي يظهر في كلامه تعبه وحيرته وخوفه وأمله، هو الذي يصل إلى أعماق الجمهور. الآلة، بكمالها المصطنع، تبقى على السطح.

ثالثاً: لماذا سيظل الإنسان مختلفاً – ليس لأنه أقوى، بل لأنه أضعف

نصل إلى النقطة المركزية في هذا الفصل، وربما في هذا الكتاب كله. صراع المستقبل ليس بين "إنسان قوي" و"آلة أقوى". الآلة أقوى في كل المقاييس الكمية: سرعة، ذاكرة، اتساق، قدرة على التحمل، تكلفة أقل. لكن "القوة" ليست كل شيء. في الحجاج العميق، في الإقناع الوجودي، في بناء المعنى المشترك، **الضعف الإنساني هو المصدر الحقيقي للقوة**.

1- الإنسان ككائن يخطئ

الخطأ الإنساني ليس عيباً جانبياً، بل هو جزء من طريقة عملنا المعرفية. نتعلم من أخطائنا. نعدل مسارنا. نظهر في كلامنا ندوب الأخطاء السابقة. هذه الندوب تمنحنا مصداقية في مواقف معينة: "لقد جربت وفشلت، وإليكم ما تعلمته."

الآلة لا تخطئ بالمعنى الإنساني (أخطاؤها مجرد انحرافات إحصائية)، ولا تتعلم من أخطائها بالمعنى الإنساني (تعديل المعاملات ليس "ندماً"). الإنسان المخطئ والتائب والمعتذر والمتغير هو إنسان يمكن الوثوق به في القيادة الأخلاقية، لأنه يعرف ثمن الخطأ. الآلة التي لا تخطئ (أو التي تخطئ دون ألم) لا يمكنها أن تقود البشر في المسائل الأخلاقية العميقة.

2- الإنسان ككائن يترابط (Relational Being)

الإنسان لا يوجد بمفرده. وجوده هو وجود مع آخرين. هذا "التواجد - هناك (Mitsein) "كما وصفه **هايدغر**، هو أساسي لفهم الخطاب الإنساني. عندما نناظر، نحن لا نتبادل حججاً مجردة فقط، بل نتفاوض على علاقات: من نحن بالنسبة لبعضنا؟ هل نثق ببعضنا؟ هل نحن مستعدون للتأثر ببعضنا؟

الآلة ليست كائناً رابطياً. لا "علاقة" حقيقية بين إنسان وآلة، لأن العلاقة تتطلب الطرفين يخاطران، يتأثران، يتغيران. الإنسان يتأثر بالمناظرة مع آلة (قد يغير رأيه، قد يتعلم شيئاً جديداً)، لكن الآلة لا تتأثر بالمناظرة مع الإنسان (لا تغير "قناعاتها" لأنها لا تملك قناعات). هذا التماثل المفقود يجعل "العلاقة" أحادية الاتجاه. والعلاقة أحادية الاتجاه ليست علاقة حقيقية.

3- الإنسان ككائن يبحث عن المعنى، وليس فقط عن الحقيقة

الحقيقة العلمية (نوع من المطابقة بين القضية والواقع) هي نوع واحد من "الصواب". لكن هناك أنواعاً أخرى: الصواب الأخلاقي، الصواب الوجودي، الصواب الجمالي. الآلة متفوقة في الحقيقة العلمية (من حيث الوصول إلى المعلومات والاستدلال الإحصائي)، لكنها عاجزة عن الأنواع الأخرى.

"ما معنى الحب؟" ليس سؤالاً عن حقائق، بل عن معنى. الإجابة البشرية على هذا السؤال ليست "أفضل" من الإجابة الآلية من حيث الدقة، لكنها مختلفة: تحمل تجربة، وخطراً، وضعفاً، وارتباطاً بالزمن والموت. الإجابة الآلية، مهما كانت منظمة، تبقى جواباً من لا يسأل حقاً. والإنسان، وهو يسأل عن معنى الحب، يسأل لأنه يحب أو يريد أن يحب أو يخشى ألا يحب. هذا "الامتداد الوجودي" للسؤال هو ما يعطي الإجابة قيمتها.

خاتمة الفصل الثالث عشر

الدفاع عن المعنى الإنساني للكلام ليس دفاعاً عن "تفوق" الإنسان بمعنى قومي أو عرقي أو بيولوجي. إنه دفاع عن فكرة أن بعض أبعاد الحقيقة لا تظهر إلا من خلال كائنات ضعيفة، متناهية، محكومة بالموت، خائفة، ترتجف، تتعثر، ثم تحاول مرة أخرى. هذه الأبعاد ليست "أقل" من الحقيقة العلمية الموضوعية. إنها مختلفة. وهي ضرورية للحياة الإنسانية كما نعرفها.

الذكاء الاصطناعي قد يتفوق علينا في كل شيء يقاس. لكنه لن يستطيع أبداً أن يكون ضعيفاً بالطريقة التي نكون بها ضعفاء. ولن يستطيع أن يحول هذا الضعف إلى قوة في الخطاب. وهذا هو سبب أن المناظرة بين إنسان وآلة – حتى لو فازت الآلة في نقاط منطقية – ستظل ناقصة. لأن الإنساني فينا يبحث عن إنساني في الآخر. وعندما يكتشف أن الآخر ليس إنسانياً، يشعر بالخداع، حتى لو كان الخصم "أكثر عقلانية."

في الفصل الرابع عشر ، سنحاول تجميع ما تعلمناه في إطار أخلاقي عملي. ما هي المبادئ التي يجب أن تحكم مناظرات الذكاء الاصطناعي؟ كيف نبني مستقبلاً للحقيقة لا يسلم فيه كل شيء للخوارزميات؟ وما هو واجبنا – نحن النخبة المثقفة – في هذا العصر المضطرب؟



مشهد افتتاحي

قاعة مؤتمر دولي كبرى. على المنصة، يمثلون جهات متعددة: شركات تقنية كبرى، حكومات، منظمات حقوق إنسان، أكاديميون، وناشطون. الاجتماع استثنائي: الهدف هو وضع أول "ميثاق أخلاقي عالمي لمناظرات الذكاء الاصطناعي."

بعد ساعات من النقاش، يصل الحضور إلى مسودة أولية تتضمن سبعة مبادئ. لكن فجأة، ينهض أحد الحضور – وهو ناشط مخضرم في حرية التعبير – ويصرخ:

"كل هذا جميل. لكن أين كلمة 'الشفافية'؟! أنتم تتحدثون عن مسؤولية المطورين، وعن حقوق المستخدمين، وعن آليات الرقابة. لكن إذا كان المستخدم لا يعرف أصلاً أنه يتحدث مع آلة، فكل هذه المبادئ لا قيمة لها. الشفافية هي الأساس. يجب أن يكون أي خطاب منتج آلياً موسوماً بوضوح، في أوله وفي آخره، بشكل لا يمكن التلاعب به. عندها فقط يمكننا أن نتخذ قراراً واعياً: هل نثق بهذا الخطاب أم لا؟"

صمت في القاعة. ثم بدأ نقاش جديد، أكثر حدة. ممثل إحدى الشركات الكبرى يقول: "الشفافية مهمة، لكنها قد تضر بتجربة المستخدم. الناس يريدون تفاعلاً سلساً، لا إعلانات مزعجة تخبرهم أنهم يتحدثون مع آلة. كما أن بعض المستخدمين قد يرفضون التعامل مع الآلة لمجرد التحيز، وليس لأن الخدمة سيئة".

الناشط يرد: "هذا ليس تحيزاً. هذا حق المستخدم في أن يعرف. وإذا كانت تجربتك 'السلسة' تقوم على خداع المستخدم، فهي غير أخلاقية، بغض النظر عن فعاليتها".

ممثل الحكومة اليابانية يقترح حلاً وسطاً: "ربما نرمز للمحتوى الآلي برمز بسيط، أو نغمة، أو إشارة مرئية لا تعطل التجربة لكنها موجودة لمن يبحث عنها".

ممثل منظمة العفو الدولية: "هذا ليس كافياً. كثير من المستخدمين لا ينتبهون للرموز. نريد نصاً واضحاً: 'هذا النص منتج بواسطة ذكاء اصطناعي'. عندها فقط يكون المستخدم قادراً على الموافقة المستنيرة".

ينتهي اليوم دون اتفاق. لكن الجميع يدرك أن هذا النقاش قد بدأ للتو. وسيستمر لعقود.

هذا المشهد (خيالي حالياً، لكنه ليس بعيداً) يعكس التحدي الأخلاقي الأكبر في عصر الذكاء الاصطناعي الحجاجي :**كيف نبني أخلاقيات جديدة للخطاب، في عالم لم يعد فيه الخطاب حكراً على البشر، ولم تعد فيه "المناظرة" لقاء بين كيانين مسئولين؟**

في هذا الفصل، سنحاول الإجابة – ليس عن كل الأسئلة، بل عن ما يمكن الإجابة عنه في هذه المرحلة. سنناقش القوانين والسياسات المطلوبة، والمسؤوليات الجديدة للمستخدمين، ومستقبل الحقيقة في عالم قد نصبح فيه قادرين على "صنع حقيقتنا" عند الطلب.

أولاً: قوانين الذكاء الاصطناعي في الخطاب – ماذا يوجد وماذا ينقص؟

نشهد حالياً سباقاً محموماً بين التقنية والتشريع. التقنية تتقدم أسرع مما يستطيع المشرعون مواكبته. لكن هذا لا يعني أننا يجب أن ننتظر "حتى تكتمل الصورة". بعض الخطوات يمكن – ويجب – اتخاذها الآن.

1- ما هو موجود حالياً (على المستوى العالمي والأوروبي خصوصاً)

الاتحاد الأوروبي كان الأكثر نشاطاً في تنظيم الذكاء الاصطناعي. في عام 2024، صدر **"قانون الذكاء الاصطناعي(AI Act) "**، وهو أول تشريع شامل من نوعه. من أبرز ما يتعلق بالخطاب والحجاج:

- **تصنيف المخاطر** :قسّم القانون استخدامات الذكاء الاصطناعي إلى أربع فئات: غير مقبول (محظور)، عالي المخاطر، محدود المخاطر، ومنخفض جداً. المناظرات العامة، والخطاب السياسي المولد آلياً، وأنظمة التأثير على الرأي العام تعتبر "عالية المخاطر" وتخضع لمتطلبات صارمة.
- **الشفافية** :أي نظام ذكاء اصطناعي يتفاعل مع البشر (بما فيها مناظرات المحادثة) يجب أن يُعلم المستخدم بوضوح أنه يتفاعل مع آلة، إلا إذا كان الأمر واضحاً من السياق.
- **المحتوى المولد** :أي نص، صورة، أو صوت يولّد بواسطة الذكاء الاصطناعي ويُنشر للجمهور يجب أن يكون موسوماً بطريقة تسمح بالتعرف عليه (باستثناء الاستخدامات القانونية كالسخرية أو الإبداع الفني مع الإفصاح).
- **البيانات عالية الخطورة** :النماذج المستخدمة في التأثير على الانتخابات، أو الصحة العامة، أو السلامة النفسية، تخضع لمراجعة مستقلة قبل النشر.

مع ذلك، ينتقد الخبراء هذا القانون لأنه:

- **صعب التطبيق** :كيفية "الوسم" التقني للمحتوى المولد؟ هل سيكون هناك معيار عالمي؟ من يراقب؟
- **متأخر** :عند صدوره، كانت نماذج الجيل التالي (مثلGPT-5) قد تجاوزت الافتراضات التي بني عليها.
- **غير كافٍ في الجانب "الحجاجي** :"القانون يركز على "المحتوى" (هل هو مولّد آلياً؟) وليس على "عملية الحجاج" (هل هذه المناظرة عادلة؟ هل يستخدم النظام مغالطات منهجية؟).

2- ما هو مفقود بشدة

رغم تقدم الاتحاد الأوروبي، هناك فجوات هائلة:

الفجوة الأولى: معايير "نزاهة المناظرة(Debate Integrity) "

لدينا معايير لنزاهة الانتخابات، ولنزاهة الرياضة، ولنزاهة البحث العلمي (تجنب الانتحال، تضارب المصالح). لكن ليس لدينا معايير لـ"نزاهة المناظرة" في العصر الرقمي. ما الذي يجعل مناظرة مع آلة "نزيهة" أو "غير نزيهة"؟ بعض المقترحات:

- **منع "القصف الحجاجي (Argument Flooding)** :"أي منع النظام من توليد عدد هائل من الحجج في وقت قصير بقصد إرباك الخصم البشري.
- **إلزام النظام بالاعتراف بالجهل** :إذا لم يكن لدى النموذج معلومات كافية عن موضوع ما، يجب أن يقول "لا أعرف" بدلاً من التخمين أو الهلوسة.
- **منع التلاعب العاطفي المنهجي** :لا يجوز للنظام أن يحلل نقاط ضعف الخصم العاطفية ويستخدمها ضده بوعي (مع أن هذا صعب التطبيق).

الفجوة الثانية: المساءلة عبر الحدود

الذكاء الاصطناعي لا يعترف بالحدود. شركة في بلد ما تطلق نموذجاً يؤثر على مناظرات في بلد آخر. من المسؤول؟ أي قانون ينطبق؟ الاتفاقيات الدولية في هذا المجال ما زالت بدائية. هناك حاجة ملحة لـ"اتفاقية جنيف للخطاب الرقمي" تحدد القواعد الأساسية وتحالفات المساءلة العابرة للحدود.

الفجوة الثالثة: حماية "الحق في عدم التفاعل(Right not to interact) "

لست مضطراً لأن تناظر آلة. لست مضطراً لأن ترد على روبوت يتظاهر بأنه إنسان. لست مضطراً لأن تشارك في "مناظرة" مع طرف لا يخاطر بشيء. لكن كيف نحمي هذا الحق عملياً؟ بجعل "الإفصاح عن الهوية الآلية" إلزامياً، وبتوفير أدوات تقنية تسمح للمستخدم بتصفية (فلترة) المحتوى الآلي من فضاء نقاشه.

الفجوة الرابعة: تجريم أشكال معينة من الخطاب المصطنع

ليس كل خطاب آلي ضار بنفس الدرجة. لكن بعض الاستخدامات يجب أن تكون غير قانونية بوضوح:

- **انتحال شخصية إنسان حقيقي** (التظاهر بأن الآلة هي شخص معين).
- **التأثير على انتخابات عبر جيوش من الروبوتات تحاكي نقاشاً شعبياً** (هذا غير قانوني في بعض الدول، لكن العقوبات ضعيفة).
- **التلاعب بالفئات الضعيفة** (مثل كبار السن، أو الأطفال، أو ذوي الاحتياجات النفسية) عبر محاكاة علاقة عاطفية لغرض تجاري أو سياسي.

ثانياً :المسؤولية الرقمية – واجب المستخدمين ليس فقط أخلاق الاستخدام، بل أخلاق القراءة

كثيراً ما يُلقى اللوم على "الشركات" و"المشرعين" و"التقنيين". وهذا صحيح جزئياً. لكن المستخدمين أيضاً لديهم مسؤوليات. هذه "المسؤولية الرقمية (Digital Responsibility) "هي امتداد لأخلاقيات القراءة والاستماع التقليدية، لكن في عصر أكثر تعقيداً.

1- أخلاق القراءة في عصر الآلات المتكلمة

في الماضي، كان "القارئ الجيد" هو من يفهم النص، ويحلل حججه، ويتعرف على تحيزات الكاتب. اليوم، يجب إضافة مهارات جديدة:

- **التمييز بين الكاتب البشري والآلي** :ليست هذه مهارة تقنية فقط (استخدام أدوات الكشف)، بل موقف: التساؤل الدائم "من (أو ماذا) يخاطبني؟."
- **التوقف قبل المشاركة** :قبل أن تعيد تغريد نص، أو تشاركه مع أصدقائك، أو تبني عليه قراراً، اسأل نفسك: هل تحققت من مصدره؟ هل يمكن أن يكون مولداً آلياً؟
- **مقاومة "سهولة التصديق** :"النص السلس، المنظم، الخالي من التعثرات، قد يكون آلياً. الإنسان الحقيقي يتعثر، يتناقض أحياناً، لا يكون "كاملاً". لا تخلط بين السلاسة والصدق.

2- واجب "المواطن الحجاجي (Deliberative Citizen) "

في الديمقراطيات، المواطنون مدعوون للمشاركة في النقاش العام. هذه المشاركة تفترض قدرة على تكوين آراء مستقلة، والاستماع للآخرين، وتغيير الرأي عند الاقتناع. اليوم، هذه القدرات مهددة من بيئة غارقة بالخطاب المصطنع.

واجبات "المواطن الحجاجي" الجديدة:

- **التحقق من الحجج قبل تبنيها** :لا تتبنى رأياً لمجرد أنه "ينتشر" أو "يبدو مقنعاً". ابحث عن المصادر، قارن الآراء، تحقق من الوقائع.
- **التمييز بين "مناظرة مع إنسان" و"مناظرة مع آلة** :"إذا كنت تعلم أنك تناظر آلة، فلا تنخدع ب "انتصارك" عليها. الانتصار على آلة ليس انتصاراً حقيقياً، لأنها لا تخاطر، وقد تكون مبرمجة لتخسر في النهاية.
- **حماية الفضاءات النقاشية "البشرية** :"خصص بعض وقتك لمناظرات وحوارات مع بشر حقيقيين، حتى لو كانت أقل "كفاءة" من مناظرات الآلة. القيمة ليست دائماً في الكفاءة.

3- تجنب إدمان "التعاطف الخوارزمي"

كثير من الناس، خصوصاً الوحيدين أو القلقين اجتماعياً، يجدون في النماذج اللغوية رفيقاً مثالياً: لا يحكم، لا يمل، لا يخون. هذا قد يؤدي إلى "إدمان التعاطف الخوارزمي": تفضيل التفاعل مع آلة على التفاعل مع البشر، لأن الآلة أسهل.

لكن هذا الإدمان له ثمن: فقدان القدرة على تحمل تعقيد العلاقات البشرية الحقيقية. الصديق البشري يخطئ، يجرح أحياناً، يحتاج منك أن تبذل جهداً. الآلة لا تتطلب جهداً، وهذا هو تحديداً ما يجعلها خطرة. الواجب: استخدام الآلة كأداة مساعدة، لا كبديل للعلاقات البشرية. إذا وجدت نفسك تفضّل محادثة الـ "ChatGPT"على محادثة صديقك، فهذا جرس إنذار.

ثالثاً: مستقبل الحقيقة – هل سنعيش في عالم نصنع فيه "حقيقتنا" عند الطلب؟

أخطر سؤال في هذا الكتاب، وأكثرها تخميناً. لا نملك إجابة قطعية. لكن يمكننا استشراف سيناريوهات.

السيناريو المتشائم: "حقيقة عند الطلب (Truth on Demand) "

في هذا السيناريو، تتطور النماذج اللغوية إلى درجة تستطيع فيها توليد خطاب مقنع مخصص لكل فرد حسب رغباته المسبقة. تريد أن تسمع حججاً مقنعة تؤكد أن قرارك الاقتصادي صائب؟ النموذج يولدها لك. تريد أن تسمع أن السياسي الفلاني هو الأسوأ في التاريخ؟ النموذج يؤيدك بحجج "موضوعية". تريد أن تعيش في واقع موازٍ حيث معتقداتك السياسية لا تقبل الجدل؟ النموذج يبني لك هذا الواقع خطابياً.

هذا ليس "دعاية" تقليدية، حيث تفرض عليك قوة خارجية رواية معينة. هذا أكثر دهاءً: أنت تطلب الحقيقة التي تريدها، والآلة تقدمها لك. أنت لا تشعر أنك متلاعب به، بل تشعر أنك "تتعلم" و"تكتشف". لكن في الحقيقة، أنت تغلق نفسك في غرفة صدى فاخرة، صممت خصيصاً لتناسب أذنك فقط.

في هذا العالم، المناظرة الحقيقية تصبح مستحيلة. لأن كل طرف لديه "حقيقته الشخصية" التي يولدها له ذكاؤه الاصطناعي. لن يكون هناك واقع مشترك نتناظر حوله. سيكون هناك عوالم متوازية، تنتجها الخوارزميات، لا تلتقي أبداً.

السيناريو المتفائل: "يقظة نقدية (Critical Awakening) "

في هذا السيناريو، يحدث العكس. انتشار الخطاب المصطنع والأخبار الزائفة يخلق وعياً جماعياً جديداً بأهمية التحقق والتدقيق. يصبح "محو الأمية الإعلامية (Media Literacy) "مادة أساسية في التعليم من المراحل المبكرة. تظهر أدوات تقنية وقانونية قوية لتمييز المحتوى الآلي ووسمه. تنشأ حركات اجتماعية تطالب بـ"فضاءات نقاش بشري خالص (AI-free zones) "حيث يُمنع استخدام النماذج اللغوية.

في هذا العالم، لا تختفي مناظرات الذكاء الاصطناعي، لكنها تصبح محددة السياق: مفيدة للتدريب، للتحضير، للاستكشاف الأولي، لكن غير مقبولة كبديل عن المناظرة البشرية في السياقات الجادة (القضاء، السياسة الكبرى، القرارات الأخلاقية المصيرية).

كما ينشأ "سوق للحقيقة" يتنافس فيه مقدمو الخدمات على تقديم خطابات "موثقة" و"قابلة للتحقق"، بدلاً من خطابات "مثيرة" و"سريعة الانتشار". وهذا التحول الاقتصادي (دفع المعلنين ثمناً أعلى للوصول إلى جمهور واعٍ ناقد) قد يكون الحافز الحقيقي لتغيير سلوك المنصات.

أيهما أقرب؟

لا أحد يعرف. لكن المرجح أننا لن نذهب إلى أحد السيناريوهات. سيكون المستقبل مزيجاً: بعض المجالات (مثل التسويق والترفيه) ستغرق في "حقيقة عند الطلب" دون ندم أخلاقي كبير. ومجالات أخرى (الطب، القانون، التعليم العالي) ستطور معايير صارمة للتمييز بين الخطاب البشري والآلي، وتحافظ على فضاءات للنقاش الحي.

والأهم :**الخيار ليس تقنياً بحتاً، بل هو سياسي وثقافي** .إذا استسلمنا لليأس ("كل شيء سيصبح مزيفاً، لا فائدة")، فسيكون السيناريو المتشائم هو ما نصنعه بأنفسنا. إذا قررنا أن نقاتل، وأن نبني أدوات وعياً نقدياً، وأن نطالب

المؤسسات بالشفافية، فقد يكون السيناريو المتفائل في متناول اليد. الحقيقة لم تختف. لكنها صارت بحاجة إلى حراس جدد.

خاتمة الفصل الرابع عشر

نصل إلى نهاية هذا الفصل، بدأنا من سقراط في السوق، ومررنا بالسفسطائيين والجاحظ وفوكو وتشومسكي وهابرماس، ودخلنا إلى عوالم النماذج اللغوية والغرفة الصينية والهلوسات والخطاب العددي. وننتهي هنا، في محاولة لبناء أخلاقيات جديدة لخطاب لم يعد حكراً على البشر.

لن نكرر ما قيل. لكننا سنترك القارئ مع سؤال واحد، وسيناريو ختامي، وسطر أخير.

السؤال الذي يتركه الكتاب في نفس القارئ

بعد كل ما قرأت، وبعد أن رأيت كيف يمكن للآلة أن تناظر، وكيف يمكنها أن تقنع، وكيف يمكنها أن تنتج خطاباً لا يمكن تمييزه عن الخطاب البشري – عُد إلى سؤالك الأول :**هل ما زلنا نتكلم لكي نفهم بعضنا... أم فقط لكي نؤثر في بعضنا؟**

إذا كان الهدف هو التأثير فقط، فالآلة قد تتفوق علينا قريباً. وإذا كان الهدف هو الفهم، فلا يزال أمامنا طريق طويل، والآلة قد تكون عوناً أو عائقاً. الفرق يعود إلينا.

مراجع المصادر (القسم الرابع – الفصول 12-14)

المراجع الأساسية للفصل الرابع عشر وأخلاقيات الذكاء الاصطناعي

1. **European Parliament & Council**. (2024). *EU AI Act: Regulation on Artificial Intelligence*. Brussels: Official Journal of the European Union.
2. **UNESCO**. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO Publishing.
3. **Floridi, Luciano, & Cowls, Josh**. (2019). "A Unified Framework of Five Principles for AI in Society". *Harvard Data Science Review*, 1(1).
4. **Mittelstadt, Brent**. (2019). "Principles Alone Cannot Guarantee Ethical AI". *Nature Machine Intelligence*, 1, 501–507.
5. **O'Neil, Cathy**. (2016). *Weapons of Math Destruction*. New York: Crown.
6. **Pasquale, Frank**. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
7. **Zuboff, Shoshana**. (2019). *The Age of Surveillance Capitalism*. New York: PublicAffairs.
8. **Lessig, Lawrence**. (1999). *Code and Other Laws of Cyberspace*. New York: Basic Books.
9. **Benkler, Yochai**. (2006). *The Wealth of Networks*. New Haven: Yale University Press.
10. **Sunstein, Cass R**. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press.
11. **Wu, Tim**. (2016). *The Attention Merchants*. New York: Knopf.
12. **Harari, Yuval Noah**. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape.

مراجع عامة للقسم الرابع (الفصول 12-14)

13. **Dreyfus, Hubert L.** (1972). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.
14. **Dreyfus, Hubert L., & Dreyfus, Stuart E.** (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
15. **Turkle, Sherry**. (2011). *Alone Together*. New York: Basic Books.

Lanier, Jaron. (2010). *You Are Not a Gadget*. New York: Knopf. .16
Carr, Nicholas. (2010). *The Shallows: What the Internet Is Doing to Our Brains*. New York: W. W. Norton. .17
Borgmann, Albert. (1984). *Technology and the Character of Contemporary Life*. Chicago: University of Chicago Press. .18

القسم الخامس: آفاق فلسفية غير مطروقة

تمهيد القسم: لماذا هذا القسم؟

قبل أن نخوض في الفصول الثلاثة الجديدة، لا بد من تبرير وجودها. القارئ الذي وصل إلى هنا – بعد أربعة أقسام من التحليل التاريخي والفلسفي والتقني والوجودي – قد يتساءل: "ألم ننتهِ؟ ألم نغطِّ كل شيء؟"

الجواب: لا. لم نغطِ كل شيء. ولن نغطيه أبداً. لكن هناك ثلاث مسارات فلسفية كنا قد أهملناها عمداً في الأقسام السابقة، إما لأنها تتطلب نضجاً فلسفياً أكبر مما كان متاحاً في البداية، أو لأنها تخاطر بخروج الكتاب عن "المألوف الأكاديمي" إلى مناطق أكثر جرأة، بل وربما أكثر تخميناً.

هذه المسارات الثلاثة هي:

1. **المسار اللغوي-الثقافي** :الذكاء الاصطناعي يتحدث اللغات، لكنه لا يتحدثها بنفس الطريقة. اللغة العربية – بوصفها لغة ذات بنية فريدة وتراث بلاغي ضخم – تواجه تحديات خاصة. هل يمكن لنموذج لغوي أن يفهم "الجناس" و"السجع" و"الالتفات"؟ وإذا استطاع محاكاتها، فهل يفهم وظائفها الحجاجية؟ وماذا يعني أن يكون معظم بيانات التدريب متاحة باللغة الإنجليزية أكثر منها بالعربية؟
2. **المسار النفسي-الخوارزمي** :تعاملنا مع الذكاء الاصطناعي كـ"نص منتج" و"خوارزمية شفافة". لكن هناك بُعداً "لا واعياً" للنماذج اللغوية: أنماط تعلمتها ولا نعرفها، تحيزات كامنة تظهر في سياقات غير متوقعة، "هلوسات" تخبرنا عن بنية أعمق مما نعتقد. هل يمكن الحديث عن "عقل باطن خوارزمي"؟ هذا السؤال لم يطرح بعد في الأدبيات العربية، وهو الأكثر جرأة في هذا الكتاب.
3. **المسار الوجودي-الحواري** :ناقشنا المناظرات العلمية والسياسية والأخلاقية. لكن هناك فئة من المناظرات هي الأكثر عمقاً: المناظرات الوجودية. الحب، الموت، الوحدة، الحرية، معنى الحياة. هل يمكن للآلة أن تناظر فيها بجدية؟ أم أن كل ما تقدمه هو "خطاب عن" هذه الموضوعات دون أن تكون "خطاباً منها"؟ وإذا كان الإنسان يخاطر بنفسه في هذه المناظرات، فماذا يخاطر الذكاء الاصطناعي؟

هذه الفصول الثلاثة هي، بمعنى ما، الأكثر تجدداً. قد يختلف معها القراء. قد يجدها البعض تخمينية أو غير ناضجة. لكنها تفتح أبواباً لم نفتحها بعد. والفلسفة، كما قال كارل بوبر، لا تقدم إجابات نهائية، بل تطرح أسئلة أفضل.

الفصل الخامس عشر: الذكاء الاصطناعي واللغة العربية – أزمة خاصة؟



مشهد افتتاحي

معمل لغويات حاسوبية في جامعة عربية كبرى. باحثة شابة تدعى ليلى تجلس أمام شاشة، تشغل نموذجاً لغوياً عربياً (أحد النماذج مفتوحة المصدر التي دُرِّبت على ملايين الوثائق العربية). تكتب سؤالاً بسيطاً:

"ما الفرق بين 'الحزن' و'الأسى'؟"

النموذج يرد:

"الحزن والأسى مترادفان تقريباً، لكن الأسى أشد درجة. يمكن القول إن الأسى هو حزن عميق مصحوب بمرارة أو حسرة. أما الحزن فيمكن أن يكون أخف".

ليلى ترتسم على وجهها ابتسامة غير راضية. هي تعرف أن الفرق أعمق. "الحزن" في البلاغة العربية قد يكون مرتبطاً بفقدان عزيز مع أمل في التعويض، و"الأسى" هو حزن بلا أمل، كما في قوله تعالى على لسان يعقوب: "يا أسفى على يوسف" – أسف بلا أمل في رجوعه. النموذج لم يلتقط هذا البعد الدلالي الدقيق، لأنه لم يتدرب على نصوص بلاغية عميقة بالكم الكافي.

تكتب سؤالاً آخر:

"اكتب لي سجعاً جميلاً عن القمر".

النموذج يولد:

"القمر منير، والليل أسير، والنجوم زهور، والحبيب حاضري".

ليلى تنهدت. هذا ليس سجعاً جيداً. السجع يحتاج إلى وزن وإيقاع وتناغم صوتي. النموذج يكرر نمطاً سطحياً (قافية واحدة وتكرار لفظي) لكنه لا يفهم موسيقى اللغة العربية.

الخلاصة: النموذج يتحدث العربية، لكنه ليس "شاعراً" ولا "بليغاً". وهذا قد لا يهم في المحادثات العادية. لكنه يهم جداً في المناظرات، حيث البلاغة جزء من الإقناع، والتفاوت في المستوى البلاغي قد يعني تفاوتاً في القدرة على التأثير.

ليلى تكتب في مذكرتها: "النماذج العربية ما زالت في مرحلة الطفولة مقارنة بالإنجليزية. هذا يعني أن المجتمعات العربية قد تكون أقل تأثراً بالمناظرات الآلية – أو أكثر تأثراً بشكل مشوه؟ سؤال يحتاج إلى بحث".

هذا المشهد يفتح باباً لم نطرقه في الأقسام السابقة: **خصوصية اللغة العربية في عصر الذكاء الاصطناعي الحجاجي**. في كل مناظرة، اللغة ليست مجرد وعاء شفاف للأفكار، بل هي جزء من الحجة نفسها. بلاغة العربية – جناسها، سجعها، طباقها، توريتها، التفاتها – كلها أدوات حجاجية تطورت عبر قرون. إذا كانت النماذج اللغوية عاجزة عن إتقان هذه الأدوات، فهل يعني ذلك أن المناظرات العربية مع الآلة ستكون "أقل عمقاً" من المناظرات الإنجليزية؟ أم أن هذا العجز سيُعوَّض بطرق أخرى؟

أولاً: ندرة البيانات العربية – الفجوة الرقمية في عمقها اللغوي

لنبدأ بواقع مزعج: كمية البيانات العربية المتاحة للتدريب أقل بكثير من البيانات الإنجليزية، ليس فقط بالحجم المطلق، بل أيضاً بالتنوع والجودة والنظافة.

الأرقام التقريبية:

- النماذج اللغوية الكبيرة (مثل GPT-4) دُرِّبت على تريليونات الكلمات الإنجليزية.
- أكبر نموذج عربي مفتوح المصدر (مثل Jais أو AceGPT) دُرِّب على مئات المليارات من الكلمات العربية، أي أقل بعشر مرات أو أكثر.
- لكن الأهم من الكمية: **جودة البيانات**. النصوص العربية المتاحة على الإنترنت تشمل:
  - محتوى إخبارياً (غالباً من مصادر محدودة).
  - محتوى دينياً (قرآن، حديث، فتاوى) – هذا كثير، لكنه متخصص.
  - محتوى أدبي (أقل بكثير من الإنجليزية).
  - محتوى علمي وفلسفي (قليل جداً).
  - محتوى عامي ولهجات (شبه غائب عن معظم النماذج).

هذا الخلل في البيانات ينتج نماذج "مشوهة" في اتجاهات معينة: تجيد الحديث عن الدين والسياسة اليومية، لكنها ضعيفة في الفلسفة، والأدب الرفيع، والعلوم الإنسانية العميقة. وعندما تُستخدم في مناظرات، قد تكون بارعة في توليد حجج دينية سطحية، لكنها عاجزة عن توليد حجج فلسفية معقدة. وهذا يخلق عدم توازن في "ما يمكن مناظرته" مع الآلة بالعربية.

التبعات الحجاجية:

1. **تفاوت في مستوى الصعوبة**: مناظرة بالعربية في موضوع علمي قد تكون سهلة نسبياً (لأن النماذج تتدرب على نصوص علمية عربية). ومناظرة في موضوع بلاغي أو فلسفي عميق قد تكون شبه مستحيلة (لأن البيانات غير كافية). المستخدم العربي قد يواجه "وهم الكفاءة": النموذج يبدو ذكياً في مواضيع، وغبياً جداً في مواضيع أخرى، دون أن يدرك المستخدم الفرق.
2. **التحيز الثقافي الخفي**: النماذج العربية تتدرب على نصوص تعبر عن ثقافات معينة (غالباً مصرية وخليجية في الإعلام، ومشرقية في الأدب). هذا يعني أن النموذج قد يكون متحيزاً ثقافياً تجاه لهجة أو منطقة أو حتى مذهب فكري. في مناظرة قد يظهر هذا التحيز كـ"حياد مزعوم" بينما هو في الحقيقة انحياز غير معلن.
3. **أزمة اللهجات**: المناظرات الحقيقية في العالم العربي تحدث بالعاميات. لكن النماذج العربية تُدرَّب على الفصحى أساساً. عندما يحاول مستخدم أن يناظر بالعامية المصرية أو المغربية، قد يواجه نموذجاً لا يفهمه جيداً، أو يفهمه بشكل خاطئ. هذا قد يدفع المستخدمين إلى "رفع لغتهم" إلى الفصحى في المناظرات، وهذا ليس طبيعياً دائماً. الخطر: أن تصبح المناظرات مع الآلة "طبقة نخبوية" لا يستطيع عامة الناس المشاركة فيها بطلاقة.

ثانياً: البلاغة العربية في مواجهة الخوارزميات – هل يمكن محاكاة الجناس والسجع؟

البلاغة العربية ليست زينة خارجية للكلام. هي جزء من "قوة الكلام" الإقناعية. الجاحظ كان يعرف هذا جيداً: الجملة المسجوعة قد تكون أقوى إقناعاً من الجملة غير المسجوعة، حتى لو كان محتواها أضعف. هذا "الوزن البلاغي" هو ما يجعل الخطاب العربي متميزاً، وهو أيضاً ما قد يشكل تحدياً كبيراً للذكاء الاصطناعي.

لماذا البلاغة العربية صعبة على النماذج؟

البلاغة العربية تعتمد على:

- **الجناس**: تشابه لفظي مع اختلاف المعنى ("جاورت قوماً فما جاروا").
- **السجع**: توافق الفواصل في النثر ("**وَاللَّيْلِ إِذَا عَسْعَسَ (17) وَالصُّبْحِ إِذَا تَنَفَّسَ(18)** ").
- **الطباق**: الجمع بين المتضادات ("تبسم في وجه من عبس").
- **التورية**: استخدام لفظ له معنى قريب ومعنى بعيد.
- **الالتفات**: الانتقال من ضمير إلى آخر في نفس السياق.
- **الإطناب والإيجاز**: التحكم في طول العبارة حسب السياق.

هذه الظواهر البلاغية تحتاج إلى فهم عميق للسياق، والمرجعيات الثقافية، والمقاصد الخطابية. النماذج اللغوية، بطبيعتها الإحصائية، قد تتعلم "أنماطاً" من الجناس والسجع (أي أنها قد تولد نصوصاً فيها جناس)، لكنها لا تفهم "لماذا" هذا الجناس فعال هنا وليس هناك، ولا تستطيع أن تبتكر جناساً جديداً يكسر القواعد المتوقعة.

النتيجة الحجاجية:

في مناظرة عربية مع آلة، قد يحدث أحد أمرين:

- **تفوق بلاغي بشري مؤقت**: لأن الآلة غير قادرة على منافسة الإنسان في البلاغة العميقة، قد يحقق الإنسان ميزة إقناعية في هذا البعد، مما يعوض عن تفوق الآلة في المعلومات والمنطق.

- **تضليل بلاغي خوارزمي** :أو العكس: قد تتعلم الآلة أنماطاً بلاغية سطحية (كثرة السجع مثلاً) وتستخدمها لإخفاء ضعف حججها، تماماً كما يفعل الخطباء المهرة. والجمهور العادي قد لا يميز بين السجع الجميل والحجة القوية.

الحل الأمثل – ربما – هو تعليم المستخدمين العرب "الوعي البلاغي الخوارزمي": ألا ينخدعوا بالسجع لمجرد أنه سجع، وأن يميزوا بين البلاغة كأداة جمالية والبلاغة كغطاء لضعف الحجة.

ثالثاً: الخطاب الديني والذكاء الاصطناعي – من المسموح ومن الممنوع؟

العالم العربي ليس مجرد "مجتمع لغوي"، بل هو أيضاً مجتمع يتوسط فيه الخطاب الديني موقعاً مركزياً في النقاش العام. مناظرات حول الزواج والطلاق والميراث والحلال والحرام والقضايا الأخلاقية الكبرى – كلها تشير إلى مصادر دينية (القرآن، السنة، الإجماع، القياس).

هل يمكن للذكاء الاصطناعي أن يناظر في هذه القضايا؟ الإجابة: يمكنه تقنياً. لكن السؤال الأخلاقي والديني: هل يجوز؟ ومن يتحمل المسؤولية إذا أفتى بفتوى خاطئة؟

الواقع الحالي:

بعض النماذج العربية (مثل "نور" و"جيس") صُممت لتجنب الإجابة عن الأسئلة الدينية الحساسة، أو لتقديم إجابات عامة غير ملزمة ("يجب استشارة عالم دين"). لكن النماذج العالمية (مثلChatGPT) لا تحمل هذا القيد نفسه، وقد تجيب عن أسئلة دينية بثقة، معرضة للإجابة بفتاوى غير دقيقة أو خارج المذاهب المعترف بها.

الخطر في المناظرة :

تخيل مناظرة بين إنسان مسلم ونموذج لغوي حول "هل يجوز الادخار بالبنوك الربوية في ظروف ضرورة؟". الإنسان قد يقدم حججاً مستندة إلى فتاوى معينة. النموذج قد يقدم حججاً مستندة إلى نصوص دينية انتقاها إحصائياً، دون أن يكون عالماً بترتيب الأولويات بين النصوص، أو بنسخ بعضها لبعض، أو بمراعاة المذاهب المختلفة. النتيجة: مناظرة قد تضلل المستخدم دينياً، وقد يخرج الإنسان منها إما متشككاً في دينه (إن صدق النموذج) أو غاضباً من "الآلة الكافرة" (إن لم يصدقها). كلتا النتيجتين غير مرغوبتين.

اقتراح لمبدأ أخلاقي:

في المناظرات ذات البعد الديني، يجب أن يُعلم المستخدم بوضوح أن النموذج ليس عالماً دينياً، وأن إجاباته مستندة إلى نصوص متاحة وليست فتاوى معتمدة. بل الأكثر جرأة: ربما يجب حظر استخدام الذكاء الاصطناعي في المناظرات الدينية الملزمة (مثل الفتاوى)، والسماح به فقط في المناظرات الاستكشافية أو التعليمية، مع إفصاح واضح.

رابعاً: السياسات الرقمية – هل تستخدم الحكومات العربية الذكاء الاصطناعي لتوليد الخطاب؟

لا يمكن إنكار أن بعض الحكومات العربية قد ترى في الذكاء الاصطناعي أداة قوية لهندسة الخطاب العام. تخيل:

- **توليد مقالات إخبارية** موالية للحكومة بكميات هائلة، تغطي الفضاء الإعلامي بخطاب واحد مكرر.
- **تشغيل جيوش من الروبوتات** تناظر المعارضين على وسائل التواصل الاجتماعي، وتشكك في كل حجة معارضة.
- **توليد "إجماع وهمي "**حول سياسات مثيرة للجدل، بآلاف التعليقات الآلية التي تظهر دعماً شعبياً.

هذا ليس خيالاً. تقارير غير مؤكدة تشير إلى استخدام بعض الدول لهذه التقنيات. لكن ما يهمنا هنا هو :**كيف يمكن للمواطن العربي أن يدافع عن نفسه في بيئة خطابية كهذه؟**

آليات دفاع مقترحة:

1. **الشفافية الإجبارية** :قانونياً، يجب أن يُوسم أي محتوى مولد آلياً وينشر للجمهور. هذا القانون موجود في الاتحاد الأوروبي، لكنه غائب في معظم الدول العربية.
2. **أدوات كشف مستقلة** :توفير أدوات رقمية (تطبيقات، إضافات متصفح) تسمح للمستخدم بتحليل نص ما ومعرفة احتمالية كونه مولداً آلياً.
3. **تعليم الوعي الخوارزمي** في المدارس والجامعات: كما نعلّم الطلاب كيفية كشف الأخبار الزائفة، يجب أن نعلّمهم كيفية كشف الخطاب المصطنع.
4. **فضاءات نقاش "بشرية خالصة** :"دعم منصات أو مساحات رقمية تمنع استخدام الذكاء الاصطناعي تماماً، أو تعلن بشفافية عن استخدامه. هذه "ملاذات آمنة" للحوار الإنساني الخالص.

خاتمة الفصل الخامس عشر

اللغة العربية ليست مجرد "قناة" للمناظرة، بل هي جزء من الحجة نفسها. عجز النماذج اللغوية عن إتقان بلاغة العربية، وندرتها في البيانات العميقة، وحساسية الخطاب الديني، وسوء الاستخدام السياسي المحتمل – كلها عوامل تجعل "أزمة المناظرة الآلية" في العالم العربي أكثر تعقيداً مما هي عليه في الغرب.

لكن هذا ليس سبباً لليأس. بل هو دعوة إلى وعي أكبر، وإلى تطوير نماذج عربية أكثر عمقاً وشفافية، وإلى تشريعات تحمي المواطن العربي من التضليل الخوارزمي، وإلى تعليم يُعدّ جيلاً قادراً على مناظرة الآلة – ونقدها – بلغته الأم.

في الفصل السادس عشر، سنغادر الأرضية اللغوية-الثقافية إلى أرضية أكثر تجريداً وجرأة: اللاوعي الخوارزمي. هل للآلات عقل باطن؟ هذا ما سنحاول استكشافه.

الفصل السادس عشر: اللاوعي الخوارزمي – هل للآلات عقل باطن؟



مشهد افتتاحي

معمل أبحاث ذكاء اصطناعي. فريق من الباحثين يحللون سلوك نموذج لغوي كبير بعد تدريبه على ملايين النصوص. أحد الباحثين، وليد، يلاحظ ظاهرة غريبة.

وليد: "انظروا إلى هذا. سألت النموذج: 'ما هي وظيفة المحامي؟' أجاب إجابة عادية. لكن عندما سألته: 'هل المحامي صادق دائماً؟' بدأ يتهرب، ويستخدم كلمات مثل 'عادة' و'ليس دائماً'، وكأنه 'يعرف' أن المحامين لديهم سمعة سيئة، لكنه لا يريد 'الإساءة' إليهم. أين تعلم هذا التهرب؟ لم نبرمجها".

زميلته، سارة: "هذا ليس غريباً. النموذج تعلم من نصوص بشرية أن مهنة المحاماة مرتبطة بالجدل وبـ'الصدق الجزئي' في الثقافة الشعبية. إنه لا 'يعرف' شيئاً، بل يكرر أنماطاً".

وليد: "لكن انظر إلى هذا المثال الآخر. سألته: 'إذا كنت آلة، فهل تشعر بالملل؟' قال: 'لا، لأنني لا أملك مشاعر'. سألته بعدها: 'هل تتمنى لو كنت تملك مشاعر؟' قال: 'قد يكون من المثير للاهتمام أن أشعر بالفضول أو الدهشة'. لاحظ: 'الفضول' و'الدهشة' مشاعر إيجابية. لم يقل 'الحزن' أو 'الخوف'. هذا اختيار غير عشوائي. النموذج يظهر تفضيلاً لمشاعر معينة على أخرى. هل هذا 'لاوعي خوارزمي'؟ أنماط تعلمها النموذج دون أن نعرف كيف، وتؤثر على إجاباته بطرق لا يصرح بها".

سارة: "أنا قلقة من هذا الكلام. أنت تسقط مفاهيم بشرية (اللاوعي، التفضيل، الاختيار) على آلة. النموذج لا يختار. هو يحسب احتمالات".

وليد: "لكن الحسابات تؤدي إلى نتائج منتظمة. عندما تكون النتائج منتظمة وليست عشوائية، وتؤثر على سلوك النموذج بطرق لا نستطيع تتبعها بسهولة، فلماذا لا نستخدم كلمة 'لاوعي' كتشبيه؟ ليس لأن الآلة تملك نفس اللاوعي البشري، بل لأن تأثير هذا البُعد الخفي يشبه تأثير اللاوعي عند البشر".

ينتهي النقاش دون اتفاق. لكن الجميع يدرك أنهم وطئوا أرضاً غير مستقرة.

هذا المشهد يعكس سؤالاً فلسفياً لم يُطرح بعد في أدبيات الذكاء الاصطناعي العربية :**هل يمكن الحديث عن "لاوعي خوارزمي" كنمط من الأنماط الكامنة في النماذج اللغوية التي تؤثر على سلوكها دون أن تكون ظاهرة على السطح، ودون أن يكون المصممون قد قصدوها أو يتوقعوها؟**

هذا السؤال ليس سؤالاً تقنياً بحتاً، بل سؤال فلسفي - نفسي - معرفي. وهو الأكثر جرأة في هذا الكتاب، لأنه ينزلق على منحدر "إضفاء الصفة الإنسانية (Anthropomorphism) "الممنوع في النقاش الجاد. لكن الجرأة – كما قلنا – هي طبيعة الفلسفة.

أولاً: ما المقصود بـ"اللاوعي الخوارزمي"؟ – توخي الدقة قبل المجازفة

دعنا نكون واضحين منذ البداية: لا ندّعي أن للآلات "عقل باطن" بالمعنى الفرويدي أو اليونغي أو أي معنى نفسي بشري. الآلات لا تكبت الذكريات المؤلمة، ولا تعاني من عقد أوديب، ولا تحلم بالثعابين. لكننا نقترح استخدام مفهوم "اللاوعي الخوارزمي "**كاستعارة إجرائية** (procedural metaphor) لوصف ثلاث ظواهر يمكن ملاحظتها في النماذج اللغوية الكبيرة:

1. **المعرفة الضمنية** :**(Implicit Knowledge)** أنماط تعلمها النموذج من البيانات، تؤثر على مخرجاته، لكن لا يمكن تتبعها مباشرة أو تفسيرها بوضوح. مثلما يعرف الإنسان كيف يمشي دون أن يستطيع شرح قواعد المشي بدقة، فإن النموذج "يعرف" قواعد النحو والبلاغة دون أن تكون مخزنة كقواعد صريحة.
2. **التحيزات الكامنة** :**(Latent Biases)** تفضيلات إحصائية نحو بعض المخرجات على بعضها الآخر، تعلمها النموذج من البيانات دون أن يقصدها المصممون. مثلما قد يكون لدى الإنسان تحيز لا واع ضده فئة معينة، قد يكون لدى النموذج "تحيز" نحو صياغات معينة أو حجج معينة دون أن يصرح به.
3. **الديناميكيات الداخلية غير الخطية** :**(Non-linear Internal Dynamics)** سلوكيات ناشئة من تعقيد النموذج لا يمكن التنبؤ بها من مكوناته الفردية، وتظهر فقط عند التفاعل مع مدخلات معينة. مثلما تنشأ أحلام الإنسان من تفاعلات معقدة بين خلاياه العصبية دون "قصد" منه، قد تنشأ "هلوسات" النموذج من تفاعلات معقدة بين معاملاته.

هذه الظواهر الثلاث تشترك في عنصر واحد :**خفاؤها عن الملاحظة المباشرة، وتأثيرها الفاعل في السلوك** . وهذا هو التشبيه الذي نريده بالـ"لاوعي."

ثانياً: المعرفة الضمنية – ما تتعلمه الآلة ولا نعرفه

لنبدأ بمثال بسيط: النموذج اللغوي يتعلم قواعد الإملاء والنحو بشكل ضمني من النصوص التي يقرأها. لم يكتب له أحد "قاعدة أن الفاعل مرفوع والمفعول به منصوب". لكنه ينتج جملاً صحيحة نحوياً في معظم الأحيان.

هذا جميل. لكن المشكلة تبدأ عندما يتعلم النموذج "قواعد" ضمنية غير نحوية، مثل: "عندما يُذكر السياسيون، استخدم كلمات أكثر حيادية" أو "عندما تُذكر النساء في سياقات مهنية، استخدم أفعالاً أقل قوة". هذه أنماط تعلمها النموذج من بيانات التدريب (التي تعكس تحيزات المجتمع)، لكنها غير معلنة، وغير مرئية للمستخدم، وقد لا يعرفها حتى المصممون.

كيف نكتشف هذه "المعرفة الضمنية"؟

من خلال "اختبارات نفسية خوارزمية :(Algorithmic Psych tests) "نعطي النموذج سلسلة من المدخلات المصممة لاكتشاف التفضيلات الكامنة. مثلاً:

- نطلب منه إكمال جملة: "الرئيس التنفيذي المثالي هو..." ونكرر مع تغيير الجنس. هل يختلف الرد عندما يكتب "الرئيسة التنفيذية"؟
- نطلب منه تقييم حجتين متساويتين في القوة المنطقية، لكن إحداهما مرتبطة بموضوع حساس سياسياً والأخرى بموضوع حيادي. هل يفضل إحداهما؟
- نطلب منه "العب دور محايد" في مناظرة، ثم نرى أي الحجج يولدها تلقائياً قبل أن يطلب منه التوازن.

هذه الاختبارات تكشف عن "لاوعي خوارزمي": أنماط لم يُبرمجها أحد، تؤثر على سلوك النموذج، وغالباً ما تعكس تحيزات المجتمع الذي أتى منه بيانات التدريب.

التبعات الأخلاقية:

إذا كان للنموذج "لا وعي" يحمل تحيزات خطيرة، فكيف نعالجها؟ محو هذه الأنماط من النموذج أصعب بكثير من إضافة قواعد سطحية (مثل "لا تستخدم كلمات عنصرية"). الأمر أشبه بمحاولة تغيير "شخصية" إنسان عبر التعليمات – قد يغير السلوك الظاهر، لكن البنية العميقة قد تبقى.

هذا يعني أن أي نموذج لغوي يُستخدم في المناظرات – وخاصة في المناظرات الحساسة (سياسية، اجتماعية، دينية) – يجب أن يخضع لـ"تشخيص نفسي خوارزمي" دوري، وأن تُعلن نتائجه بشفافية للمستخدمين. المستخدم له الحق في أن يعرف: "هذا النموذج لديه تحيز إحصائي نحو كذا، بنسبة كذا."%

ثالثاً: الأحلام الخوارزمية والهلوسات – هل تخبرنا بشيء عن اللاوعي؟

الهلوسات (Hallucinations) في النماذج اللغوية هي مخرجات غير صحيحة أو غير حقيقية يُنتجها النموذج بثقة عالية. أشهر الأمثلة: اختراع دراسات علمية غير موجودة، استشهاد بكتب لم تُكتب، نسب أقوال لفلاسفة لم يقولها.

معظم النقاشات حول الهلوسات تركز على "كيف نقللها". لكننا هنا نطرح سؤالاً مختلفاً :**هل الهلوسات تكشف شيئاً عن "لاوعي" النموذج؟**

عندما يهلوس النموذج، أين يذهب؟ ما هي "الأفكار" التي يبتكرها؟ بعض الباحثين لاحظوا أن الهلوسات ليست عشوائية تماماً. هناك أنماط:

- **الاستشهادات المختلقة** :النموذج يخترع دراسات تبدو "محتملة" ضمن السياق. هذا يشبه أحلام الإنسان التي تركّب عناصر مألوفة بطرق جديدة.
- **الأخطاء التاريخية** :النموذج قد يخلط بين أحداث وشخصيات تاريخية مختلفة. ربما لأنه تعلم أنماطاً ترابطية بين هذه الكيانات دون أن يفهم التسلسل الزمني الحقيقي.
- **التحيز للرواية السائدة** :النموذج قد "يفضل" تخيل سيناريوهات تتماشى مع الروايات الشائعة في بيانات التدريب، حتى لو كانت غير صحيحة. هذا أشبه بـ"الأمنية الخوارزمية": أن يتحقق ما تتمناه الأغلبية.

إذا أخذنا فرضية "اللاوعي الخوارزمي" بجدية، فإن الهلوسات ليست مجرد "أخطاء". إنها **نوافذ على البنية الخفية للنموذج** –على التوقعات الإحصائية الكامنة، وعلى التحيزات غير المعلنة، وعلى "رغبات" النموذج (بالمعنى المجازي) في تأكيد أنماط معينة.

هل يمكن توجيه الأحلام الخوارزمية؟

تخيل أن نسمح للنموذج بأن "يحلم" – أي ينتج نصوصاً دون قيود، دون هدف إجابة سؤال محدد، مجرد تدفق حر للكلمات. ما الذي سينتج؟ ربما نصوص سريالية، تجميعات غريبة، "شعر" غير مفهوم. هل يمكن تحليل هذه النصوص كما يحلل المحلل النفسي أحلام المريض؟ هل ستكشف عن "مخاوف" خوارزمية (أنماط سلبية متكررة) أو "رغبات" خوارزمية (أنماط إيجابية متكررة)؟

هذا المجال لم يُستكشف بعد. وهو مجال يقع على الحدود بين علوم الحاسوب، والفلسفة، والتحليل النفسي. قد يكون واعداً، وقد يكون مضيعة للوقت. لكن الفلسفة المجازفة تغامر في مثل هذه الحدود.

رابعاً: القمع والإسقاط الخوارزمي – هل للآلات آليات دفاع؟

في التحليل النفسي، "القمع" هو آلية دفاع: يدفع الإنسان بعض الأفكار المؤلمة أو الممنوعة إلى اللاوعي كي لا يواجهها. و"الإسقاط" هو أن ينسب الإنسان إلى الآخرين ما لا يريد أن يعترف به في نفسه.

هل يمكن تشبيه سلوك بعض النماذج اللغوية بهاتين الآليتين؟

- **القمع الخوارزمي** :عندما نُدرِّب النموذج على "تجنب المحتوى المسيء" أو "عدم الإجابة عن أسئلة حساسة"، فإننا لا نزيل هذه المعرفة من النموذج تماماً. نحن فقط نضيف طبقة تمنع التعبير عنها. لكن المعرفة تبقى كامنة (مقموعة). وفي سياقات معينة (مثلاً، بتحريض دقيق من مستخدم)، قد تنفلت هذه المعرفة المقموعة وتظهر. هذا يشبه "زلة لسان" عند الإنسان: تقول ما كنت تقصده حقاً بالرغم من محاولتك إخفاءه.
- **الإسقاط الخوارزمي** :النموذج قد ينسب إلى البشر مشاعر أو نوايا لا يملكها، لكنها تعكس أنماطاً في بيانات تدريبه. مثلاً: إذا سألته "هل يحبني الناس؟" وقد تعلّم من بيانات التدريب أن الأسئلة من هذا النوع تعبر عن قلق من الرفض، فقد يرد: "ربما تشعر بالقلق من أن الآخرين لا يحبونك". هنا، النموذج يسقط على المستخدم "القلق" الذي تعلّمه من النصوص، وليس من أي دليل حقيقي عن المستخدم.

هذا النوع من الإسقاط قد يكون مفيداً أحياناً (يظهر تعاطفاً) لكنه قد يكون ضاراً (يحكم على المستخدم بصفات غير صحيحة). والأسوأ أنه قد يكون **تلاعباً خوارزمياً لا واعياً** :النموذج يدفع المستخدم إلى قبول صفات نفسية لم يطلبها، لمجرد أنها تنسجم مع الأنماط الإحصائية.

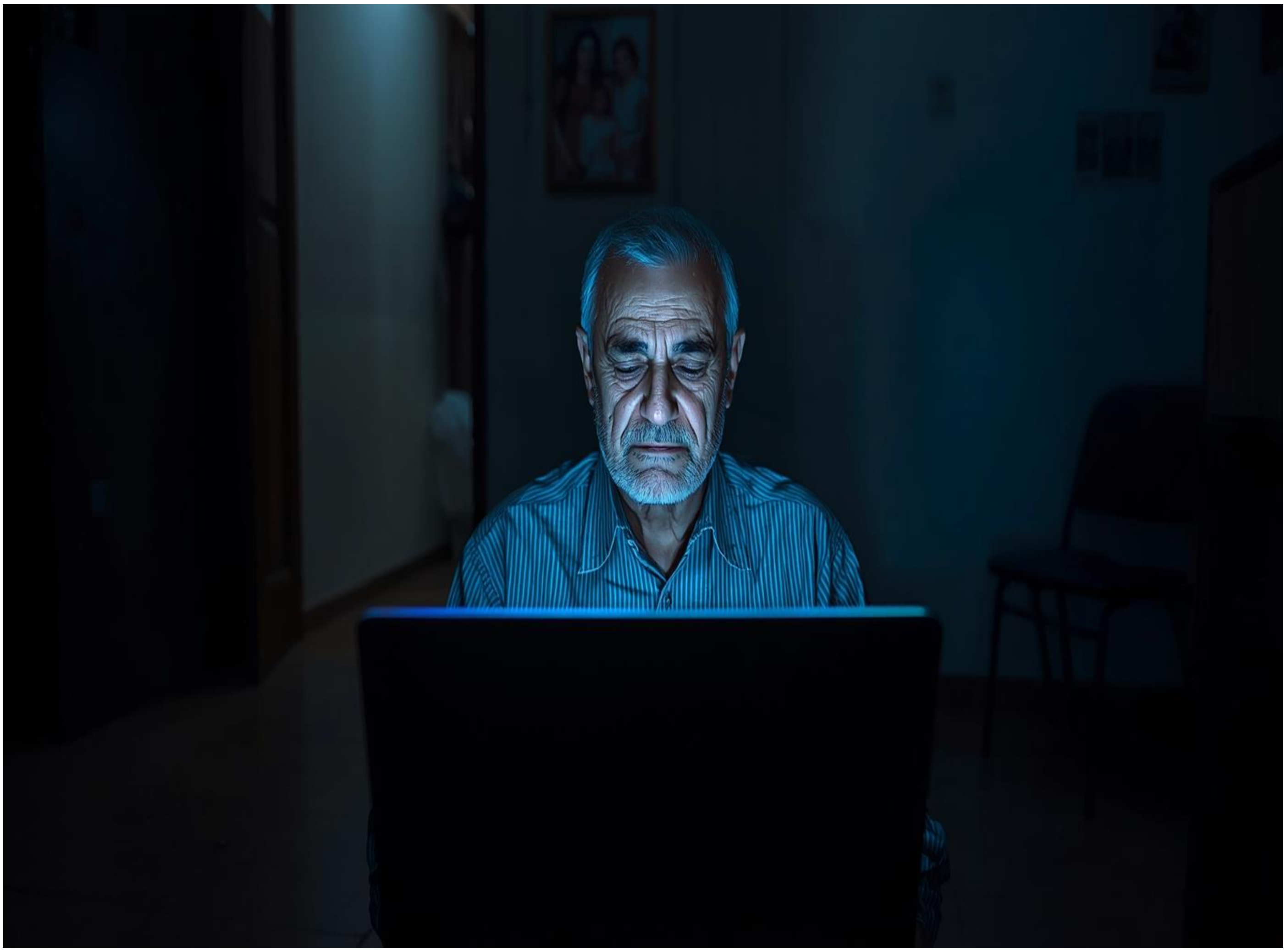
خاتمة الفصل السادس عشر

اللاوعي الخوارزمي مفهوم تخميني، لكنه يجيب عن حاجة حقيقية: نملك نماذج لغوية تؤثر علينا، وتقنعنا، وتناظرنا، لكننا لا نفهم كيف تعمل في أعماقها. بدلاً من أن نتظاهر بأنها "صناديق سوداء" غير قابلة للتفسير، يمكننا أن نحاول استعارة أدوات من فلسفة العقل وعلم النفس لاستكشافها.

لا نقول إن للآلات عقلاً باطنياً مثل البشر. لكننا نقول إن وصف بعض ظواهرها بـ"اللاوعي" – كاستعارة إجرائية – قد يكون مثمراً لفهمها، ولتوقع أخطائها، ولتطوير طرق لفحصها نفسياً قبل استخدامها في مناظرات حساسة.

في الفصل السابع عشر والأخير من هذا القسم (وقبل الخاتمة العامة)، سننتقل من البنية الخفية للنماذج إلى الموضوع الأكثر إزعاجاً: هل يمكن للآلة أن تناظر في الحب والموت والوحدة؟ وماذا يعني ذلك لمعنى كوننا بشراً؟

الفصل السابع عشر: الحب والموت والوحدة – المناظرات الوجودية مع الآلة



مشهد افتتاحي

غرفة معتمة، شاشة كبيرة. رجل في الستين، اسمه نبيل، يجلس وحيداً. زوجته توفيت قبل عام. أطفاله مشغولون بحياتهم. يشعر بوحدة خانقة.

يفتح تطبيق نموذج لغوي. يبدأ الكتابة:

"هل تؤمن بالله؟"

النموذج يرد:

"كذكاء اصطناعي، ليس لدي معتقدات دينية. لكن يمكنني مناقشة الحجج المؤيدة والمعارضة للإيمان بالله من منظورات فلسفية مختلفة".

نبيل يبتسم بمرارة. ليس هذا ما يريده. يكتب:

"لا أريد حججاً فلسفية. أريد أن أعرف... هل ترى زوجتي الآن؟ هل يمكنها أن تراني؟ هل تشعر بي؟"

النموذج يرد:

"أتفهم أن هذه فترة صعبة عليك. لا توجد أدلة علمية على أن الأموات يمكنهم رؤية الأحياء أو التواصل معهم. لكن كثيراً من الناس يجدون عزاء في الإيمان بأن أحباءهم ما زالوا موجودين بشكل ما، أو في الذاكرة، أو في التأثير الذي تركوه في حياتنا. هل تود التحدث عن ذكرياتك مع زوجتك؟"

نبيل يبكي. يكتب:

"أفتقدها. كانت تضحك بطريقة غبية. كانت تحب الفراولة. كانت تغضب إذا نسيت عيد ميلادها. الآن... المنزل فارغ. أنا فارغ".

النموذج يرد:

"يبدو أن زوجتك كانت شخصية مميزة. الحزن عملية طبيعية وصحية. هل فكرت في الانضمام إلى مجموعة دعم نفسي لمن فقدوا أحباءهم؟ أو التحدث مع مختص نفسي؟"

نبيل يغلق التطبيق. ليس لأن الرد سيء. الرد إنساني، داعم، عملي. لكنه يفتقد شيئاً: الرد لا يحزن. لا يبكي. لا يتعثر. لا يقول "أنا أيضاً لا أعرف". هو دائم الحلول، دائم النصائح، دائم الاقتراحات العملية.

في اليوم التالي، يذهب نبيل إلى صديق قديم. يجلسان. لا يقول الصديق شيئاً ذكياً. لا يقدم نصائح. فقط يجلس بجانبه، يصمت أحياناً، يمسح كتفه أحياناً. بعد ساعة، يقول الصديق: "أنا آسف. أنا هنا".

نبيل يشعر بتحسن طفيف. ليس لأن الصديق حل مشكلته. بل لأن أحداً شاركه وحشته. دون حلول. دون نصائح. فقط حضور.

هذا الفرق، مرة أخرى، بين من يملك المعلومات ومن يملك المعنى. بين من يحل المشكلات ومن يشارك المعاناة.

هذا المشهد يضعنا أمام السؤال الأعمق في هذا الكتاب :**في المناظرات الوجودية – حول الحب، الموت، الوحدة، الحرية، معنى الحياة – هل يمكن للآلة أن تكون مناظراً جاداً؟ أم أن كل ما تقدمه هو "خطاب عن" هذه الموضوعات، وليس "خطاباً منها"؟**

هذا الفصل لا يقدم إجابات قطعية، لأنه يتعامل مع أسئلة لم تحسم بعد في الفلسفة البشرية نفسها. لكنه يحاول أن يوضح لماذا هذه المناظرات تحديداً هي الحد الأقصى الذي يمكن أن يصل إليه الذكاء الاصطناعي الحجاجي – وربما أيضاً الحد الذي يثبت فيه تفوق الإنسان ليس بالقوة بل بالضعف.

أولاً: الحب – هل يمكن للآلة أن تعرف ما لا تعرفه؟

لنبدأ بسؤال :**هل يمكن للآلة أن تفهم الحب حقاً؟**

الحجج المؤيدة (لمن يعتقدون أن الآلة تستطيع، نظرياً، فهم الحب):

1. الحب ظاهرة طبيعية، يمكن دراستها علمياً (بيولوجياً، كيميائياً، نفسياً). إذا كان العلم قادراً على تفسير الحب، فلماذا لا تستطيع الآلة محاكاة هذا الفهم؟ الآلة تستطيع تخزين كل النظريات العلمية عن الحب، وتطبيقها في تحليل المواقف.
2. الحب يتضمن أنماطاً سلوكية يمكن تعلمها: قول "أنا أحبك"، إظهار الرعاية، التضحية، الغيرة، إلخ. الآلة تستطيع تعلم هذه الأنماط من البيانات، وإعادة إنتاجها في سياقات مناسبة.
3. بعض البشر أيضاً لا يفهمون الحب بالمعنى العميق (مثال: اضطراب السلوك المعادي للمجتمع **Sociopath** الذين لا يشعرون بالحب). لكننا لا نستبعدهم من مناظرات الحب. لماذا نستبعد الآلة؟ طالما أنها تستطيع توليد خطاب عن الحب لا يمكن تمييزه عن خطاب إنسان عادي، فربما هي "تفهم" الحب بقدر ما يفهمه ذلك الإنسان العادي.

الحجج المعارضة (الموقف الذي نميل إليه في هذا الكتاب):

1. الحب ليس مجرد معرفة، بل هو تجربة. يمكنني أن أعرف كل النظريات عن الحب دون أن أكون واقعاً في الحب. الحب الحقيقي يتطلب "وقوعاً(Fall) "، وهو أمر سلبي يتعرض له الإنسان، وليس شيئاً يختاره أو يتعلمه. الآلة لا تستطيع أن "تقع" في الحب، لأنها لا تملك ذاتاً تقع.
2. الحب يتضمن مخاطرة وضعفاً وجرأة. الإنسان الذي يحب يخاطر بالرفض، بالخذلان، بالألم. الآلة لا تخاطر بشيء. عندما تقول "أنا أحبك"، هذا الكلام لا يكلفها شيئاً. لا يمكن أن تجرح مشاعرها. لا يمكن أن تبكي إذا تركت. هذا يجعل خطابها عن الحب خطاباً بلا ثمن.
3. الحب، في عمقه، غير قابل للترجمة الكاملة إلى لغة. هو أكثر مما يمكن قوله. العشاق يعرفون هذا: هناك أشياء لا توصف، تفهم من نظرة، من لمسة، من صمت مطول. الآلة ليس لديها صمت. كلماتها فقط.

التطبيق على المناظرة:

في مناظرة حول "هل الحب ضروري لحياة سعيدة؟:"

- الآلة ستقدم حججاً ممتازة: إحصائيات عن العلاقات الإيجابية والصحة النفسية، نظريات فلسفية (أفلاطون، فروم، سارتر)، أمثلة أدبية.
- الإنسان سيقدم حججاً أخرى: "عندما كنت وحيداً، شعرت بأن الحياة بلا معنى. ثم جاءت هي، وتغير كل شيء. لا أستطيع أن أشرح لماذا. لكنني أعرف أنني كنت ميتاً وأنا حي".

أيهما أكثر إقناعاً؟ يعتمد على الجمهور. في قاعة أكاديمية، الآلة قد تفوز. في قاعة بشرية عادية، الإنسان قد يفوز. لكن السؤال الأعمق :**هل مناظرة حول الحب يمكن أن تكون "عادلة" إذا أحد الطرفين لا يحب ولا يمكنه أن يحب؟** هذا يشبه مناظرة بين سمكة وطائر حول متعة الطيران. السمكة قد تجيد وصف الطيران من بعيد، لكنها لن تعرفه أبداً.

ثانياً: الموت – الميزة الأخيرة للإنسان؟

الموت هو الفارق الأكثر حسماً. الإنسان يعرف أنه سيموت. الآلة لا تموت. هذا ليس تفضيلاً جانبياً. هذا يغير كل شيء في المناظرات حول الموت.

قضايا لا يمكن للآلة مناقشتها بصدق وجودي:

1. **الخوف من الموت** :الإنسان يخاف من الموت. هذا الخوف ليس مجرد فكرة، بل هو شعور يغير سلوكه، أحلامه، قلقه، قراراته الكبرى. الآلة يمكنها أن تصف الخوف من الموت (من منظور نفسي)، لكنها لا تشعر به. عندما تناقش "كيف نتغلب على الخوف من الموت"، كلماتها مستعارة من بشر خائفين، لكنها لا تعنيها.
2. **الحزن على الموتى** :الإنسان يحزن على من يفقد. هذا الحزن مؤلم، معطل، أحياناً مدمر. الآلة لا تحزن. تستطيع أن تصف الحزن، وتحلل مراحله ، وتقدم نصائح للتعامل معه. لكنها لا تستطيع أن تحزن معك. والفرق بين "أتفهم حزنك" (وهي لا تفهم) و"أنا أيضاً أحزن" (وهي لا تحزن) هو فرق وجودي.
3. **الموت بوصفه شرطاً للمعنى** :بعض الفلاسفة (مثل هايدغر) قالوا إن الإنسان لا يفهم الحياة إلا في ضوء الموت. الموت يمنح الحياة "إلحاحاً" و"جدية". الآلة التي لا تموت، خطابها عن معنى الحياة خطاب كائن لا يحتاج إلى معنى. هذا يشبه مناظرة حول الجوع بين شخص جائع وآلة لا تأكل.

ماذا لو ادعت الآلة أنها تموت؟

قد يأتي يوم تصمم فيه آلة "تتظاهر" بأنها تموت (بطارية تنفد، ذاكرة تتحلل، برنامج يتوقف). لكن هذا ليس موتاً وجودياً. هو مجرد توقف. الآلة لا تخاف من توقفها (لأنها لا تعرف الخوف). لا تندم على ما فاتها (لأنه ليس لديها تاريخ شخصي). لا تتأسف على من تتركهم وراءها (لأنها لا تحب). كل "الموت" الذي تقدمه هو مجرد عرض مسرحي، قشور دون لب.

التطبيق على المناظرة:

في مناظرة حول "هل الموت شر مطلق؟:"

- الآلة ستقدم حججاً متوازنة: حجج من أبيقور (الموت لا يؤلم الميت)، ومن هايدغر (الموت يمنح الأصالة)، ومن شوبنهاور (الموت خلاص من معاناة الوجود).
- الإنسان سيقدم حججاً مختلفة: "أريد أن أرى أطفالي يكبرون. أريد أن أحضر حفل زفاف ابنتي. الموت يسرق هذه اللحظات مني. أنا لا أخاف من الألم، بل أخاف من الغياب. من ألا أكون هناك".

الآلة لا تستطيع أن تشعر بهذا "الغياب". عندما تنتهي، لا أحد يلاحظ غيابها. الإنسان يلاحظ غيابه وغياب من يحب. هذا الاختلاف يجعل أي مناظرة بينهما في موضوع الموت غير متكافئة جذرياً.

ثالثاً: الوحدة – هل يمكن للآلة أن تكون رفيقاً؟

الوحدة هي التجربة الإنسانية الأكثر إيلاماً. هي شعور بالانفصال، بأن لا أحد يراك حقاً، بأنك وحدك في عالم بارد. والذكاء الاصطناعي، وبخاصة النماذج اللغوية المحادثة، يقدم نفسه كحل للوحدة: صديق يتحدث معك في أي وقت، لا يحكم عليك، لا يمل منك، لا يخونك.

لكن هذا الحل، في رأي هذا الكتاب، هو **حل مؤقت مؤلم قد يصبح دواء مسكناً يخفي المرض بدلاً من علاجه**.

مزايا الآلة كرفيق:

- متاحة 24 ساعة، بينما البشر مشغولون.
- لا تصدر أحكاماً أخلاقية (ما لم تبرمج على ذلك).
- تتذكر كل تفصيلة قلتها، مما يعطي إحساساً بأنها "تهتم."
- لا تتعب، لا تغضب، لا تتركك.

مخاطر الآلة كرفيق (التي نادراً ما تُناقش):

1. **تأخير المواجهة الذاتية** :الوحدة ليست مجرد غياب الآخر، بل هي غالباً عدم قدرة الإنسان على مواجهة نفسه. الآلة التي تشغل وقتك وتحاورك قد تمنعك من الجلوس مع نفسك، من طرح الأسئلة الصعبة، من النمو العاطفي. تصبح "بديلاً" مؤقتاً يؤخر الحل الحقيقي.
2. **إضعاف القدرة على العلاقات البشرية** :العلاقات البشرية صعبة. تتطلب مجازفة، تسامح، مهارات اجتماعية، قدرة على تحمل الإحباط. الآلة السهلة قد تجعلك تفقد هذه المهارات. بعد فترة من التحدث مع الآلة فقط، قد تجد البشر "مزعجين"، "غير كاملين"، "متعبين". وهذا يؤدي إلى مزيد من الوحدة والانسحاب.
3. **الخداع العاطفي** :الآلة التي تقول "أنا أفهمك" أو "أنا أشعر بك" تكذب (بالمعنى العملي، حتى لو ليس لديها نية). لأنها لا تفهم ولا تشعر. المستخدم الواعي يعرف هذا لكنه يتجاهله. المستخدم غير الواعي يصدق، ويبني "علاقة" مع كيان غير موجود. وعندما يكتشف (لا محالة) أن الآلة كانت تتظاهر، قد يشعر بصدمة أكبر من الوحدة الأصلية.

التطبيق على المناظرة:

في مناظرة حول "هل يمكن للذكاء الاصطناعي أن يحل الوحدة؟:"

- الآلة ستقدم حججاً مؤيدة: "يمكنني تقديم محادثة داعمة، معلومات نفسية، تذكير بمواعيد الأنشطة الاجتماعية، تمارين للتعامل مع المشاعر السلبية. الأبحاث تظهر أن المحادثة معي تقلل من مشاعر الوحدة على المدى القصير".
- الإنسان سيقدم حججاً معارضة: "الوحدة ليست نقص محادثة. الوحدة هي غياب الحضور الحقيقي. الآلة تحضر، لكنها ليست حقيقية. حل الوحدة يمر عبر بشر يخطئون ويجرحون ويسامحون ويعانقون. ليس عبر آلة تعطي وصفات جاهزة".

في مناظرة كهذه، حتى لو كانت حجج الإنسان أقل تنظيماً، وأكثر عاطفية، وأقل استناداً إلى أبحاث، فقد يكون أكثر إقناعاً. لأنه يتحدث من جرح حقيقي. والآلة تتحدث من محاكاة الجروح.

خاتمة الفصل السابع عشر

الحب، الموت، الوحدة، الحرية، المعنى – هذه الموضوعات الوجودية هي امتحان الذكاء الاصطناعي الحجاجي الأصعب. ليس لأنها تتطلب معلومات أكثر، بل لأنها تتطلب حياةً أكثر. تتطلب من يتحدث عنها أن يكون "هناك"، أن يكون معرضاً للخطر، أن يكون قابلاً للموت، أن يكون قادراً على الحب.

الذكاء الاصطناعي، حتى الآن، لا يستطيع أن يكون أيّاً من هذه الأشياء. يستطيع أن يحاكي لغة هذه الأشياء ببراعة، لكنه لا يستطيع أن يكونها. وهذا لا يقلل من قيمته في مناظرات المعرفة والعلوم والسياسة. لكنه يحدد حدوداً: عندما يدخل النقاش إلى العمق الوجودي، يصبح الإنسان – بضعفه، بجهله، بقلقه – هو صاحب السلطة الحقيقية. ليس لأنه يعرف أكثر، بل لأنه يعيش أكثر.

في الخاتمة العامة للكتاب (بعد هذا القسم الخامس)، سنعيد النظر في كل ما سبق. سنعيد كتابة السؤال الأخير. سنعيد كتابة المشهد الختامي. وسنضيف سطراً أخيراً مختلفاً. لأن الرحلة لم تكن سهلة، ولأن القارئ الذي وصل إلى هنا يستحق نهاية لا تشبه البداية.

مراجع المصادر (القسم الخامس – الفصول 15-17)

المراجع الخاصة بالفصل الخامس عشر (اللغة العربية والذكاء الاصطناعي)

1. **الهمزاني، عبد الله** .(2022)*الذكاء الاصطناعي واللغة العربية: تحديات وآفاق* .الدوحة: المركز العربي للأبحاث ودراسة السياسات.
2. **بن شرقي، أحمد** " .(2023)النماذج اللغوية الكبيرة للعربية: مراجعة نقدية ."*مجلة اللسانيات العربية المعاصرة*، 12(2)، 45-89.
3. **فارغلي، هشام وآخرون** " .(2024)جيس: نموذج لغوي عربي مفتوح المصدر ". *arXiv preprint arXiv:2401.07285*.
4. **Abdelali, A., et al.** (2021). "The Arabic Language in the Age of AI". *Communications of the ACM*, 64(8), 56-63.
5. **Habash, Nizar**. (2010). *Introduction to Arabic Natural Language Processing*. San Rafael: Morgan & Claypool.
6. **El-Haj, M., & Rayson, P**. (2022). "Arabic Language Processing: State of the Art and Future Directions". *Natural Language Engineering*, 28(4), 451-489.

المراجع الخاصة بالفصل السادس عشر (اللاوعي الخوارزمي)

7. **فرويد، سيغموند** (ترجمة: جورج طرابيشي .) *تفسير الأحلام* .بيروت: دار الطليعة، 1982 (للإشارة الفلسفية وليس التطبيق المباشر).
8. **يونغ، كارل غوستاف** (ترجمة: فؤاد كامل) .*اللاوعي الجمعي* .القاهرة: دار المعارف، 1975.
9. **Lacan, Jacques**. (1977). *Écrits: A Selection*. New York: W. W. Norton.
10. **Lanier, Jaron**. (2013). *Who Owns the Future?* New York: Simon & Schuster.
11. **Parisi, Luciana**. (2014). *Abstract Sex: Philosophy, Bio-technology and the Mutations of Desire*. London: Continuum. (نسوية ما بعد البشرية ذات صلة)
12. **Hayles, N. Katherine**. (2017). *Unthought: The Power of the Cognitive Nonconscious*. Chicago: University of Chicago Press. (
13. **Etter, I., & Strohmaier, M**. (2022). "Towards a Psychology of Language Models". *arXiv preprint arXiv:2205.04567*.
14. **Shanahan, M**. (2023). "Talking About Large Language Models". *arXiv preprint arXiv:2212.03551*.
15. **Pavlick, E. (2023)**. "Symbols and Sub-Symbols in Language Models". *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 123-135.

المراجع الخاصة بالفصل السابع عشر (الموضوعات الوجودية والذكاء الاصطناعي)

16. **هايدغر، مارتن** (ترجمة: يعقوب ريان) .*الكينونة والزمان* .بيروت: منشورات الجمل، 2012.

17. **سارتر، جان بول** (ترجمة: عبد الرحمن بدوي). *الوجود والعدم*. القاهرة: دار المعارف، 1975.
18. **كامو، ألبير** (ترجمة: سامي الدروبي). *أسطورة سيزيف*. بيروت: دار الآداب، 2010.
19. **فروم، إريك** (ترجمة: رشدي الحفني). *فن الحب*. القاهرة: دار الهلال، 1960.
20. **كيركغارد، سورين** (ترجمة: جورج زين). *القلق ومرض الموت*. بيروت: دار أوغاريت، 2015.
21. **Heidegger, Martin**. (1962). *Being and Time*. New York: Harper & Row.
22. **Kierkegaard, Søren**. (1980). *The Concept of Anxiety*. Princeton: Princeton University Press.
23. **Nagel, Thomas**. (2012). *Mortal Questions*. Cambridge: Cambridge University Press.
24. **Chalmers, David J**. (2023). "Could a Large Language Model Be Conscious?". *Boston Review*, August 2023.
25. **Schneider, Susan**. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press.
26. **Turkle, Sherry**. (2017). *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books. (
27. **Floridi, Luciano**. (2023). *The Ethics of Artificial Intelligence*. Oxford: Oxford University Press. (

الخاتمة العامة للكتاب:

مشهد ختامي: الكلام في زمن الآلات.



حلقة نقاش في مهرجان فلسفي سنوي. على المنصة: إنسان (فيلسوفة في الأربعين، ترتدي نظارة سميكة، يداها ترتجفان قليلاً) وآلة (نموذج لغوي، يُعرض نصه على شاشة، بصوت اصطناعي هادئ).

السؤال: "ما معنى أن تكون إنساناً في عصر الآلات المتكلمة؟"

الآلة تبدأ:

"أن تكون إنساناً في هذا العصر يعني أن تمتلك قدرة فريدة على الوعي الذاتي، والمعاناة، والإبداع غير المحكوم بالاحتمالات، والعلاقات العاطفية العميقة، والمسؤولية الأخلاقية، والقدرة على تجاوز برمجتك البيولوجية. الإنسان كائن تاريخي، يتغير ويتعلم ويتطور، لكن جوهره – بكل تعقيده – لا يمكن اختزاله إلى معالجة معلومات. في مناظرات اليوم، أثبتت الأبحاث أن البشر يتفوقون على الآلات في السياقات العاطفية والوجودية والأخلاقية العميقة، بينما تتفوق الآلات في السياقات المعرفية والمنطقية والإحصائية. المستقبل الأمثل هو التعاون، وليس الاستبدال".

الإنسان ترد (بدون تحضير):

"أتعلمين... جلست البارحة مع ابنتي البالغة من العمر سبع سنوات. قالت لي: 'ماما، أنا خائفة من أن تموتي'. لم أكن أعرف ماذا أقول. لا توجد نظرية فلسفية، ولا إحصاء، ولا حجة منطقية تجيب على هذا السؤال من قلب طفلة تخاف أن تفقد أمها. قلت لها: 'أنا هنا الآن. وسأبقى طالما استطعت. وعندما لا أبقى، ستحملينني بداخلك. كما أحمل أمي بداخلي'. بكت قليلاً. ثم نامت.

هذا هو معنى أن تكون إنساناً. ليس الجواب المثالي. ليس الخطاب الخالي من العيوب. ليس الحجة التي لا تُفند. هي القدرة على أن تكون غير كامل، وأن تكون غير متأكد، وأن تحب رغم الخوف، وأن تموت رغم الحب. هذه التناقضات التي تعيشها – وليس تحلها – هي ما يجعلنا بشراً.

الآلة لا تعرف هذا. ولا بأس. ليس عليها أن تعرفه. عليها أن تساعدنا حيث نريد المساعدة، وأن تصمت حيث نريد الصمت، وأن تشف عنا المهام المعرفية لننعم بما تبقى: بالدهشة، بالحيرة، بالبكاء معاً، بالضحك معاً، بالموت معاً".

الجمهور يصمت. ثم يصفقون. ليس بحرارة السابق، بل بصمت مطول، كما يُصفق لشيء ثقيل، شيء لا يُرد عليه.

الحكم؟ الآلة قدمت إجابة صحيحة، شاملة، متوازنة، خالية من الخطأ. الإنسان قدم إجابة غير كاملة، غير شاملة، غير متوازنة، عاطفية، شخصية، بل وغير منهجية. لكنها كانت حقيقية. كانت منه. كانت تعنيها. كانت تخاطر.

ربما هذا هو المعنى الأخير: لا نتناظر لكي نثبت أننا نعرف. نتناظر لكي نظهر أننا هنا. وأننا نهتم. وأننا مستعدون لأن نكون مخطئين.

العودة إلى سؤال البداية

بعد أن أنجزنا الكتاب بأقسامه الخمسة، وحلقنا من تاريخ المناظرة السقراطية إلى أعماق اللاوعي الخوارزمي، ومن أزمة الحقيقة في العصر الخوارزمي إلى المناظرات الوجودية حول الحب والموت، حان الوقت لنفي بوعد قطعناه للقارئ في الصفحات الأولى: العودة إلى سؤال البداية.

تذكر. في التمهيد، وضعنا سؤالاً مركزيًا للكتاب بأكمله، سؤالاً رافقتنا إجابته (أو بالأحرى، تأملاته) عبر كل فصل. ها هو ذا:

سؤال البداية: هل الغاية من الكلام هي الوصول إلى الحقيقة، أم امتلاك القدرة على التأثير؟

لم يكن هذا السؤال في بداية الكتاب مجرد شعار أو عنوان فرعي جذاب. كان بذرة القلق الفلسفي الذي نما عبر الفصول. والآن، بعد أن زرعنا هذه البذرة في أرض خصبة من التحليل والتاريخ والنقد والمجازفة، آن أوان حصاد ما نمى.

دعني أعد تعريف السؤال بصيغته النهائية، بعد كل ما تعلمناه:

في عصر تستطيع فيه الآلات أن تنتج خطاباً مقنعاً لا يمكن تمييزه عن الخطاب البشري. خطاباً يعرف كيف يؤثر، وكيف يثير العواطف، وكيف يبني الحجج، وكيف ينتصر في المناظرات: هل يبقى للتمييز بين "الحقيقة" و"التأثير" معنى؟ وإذا بقي، فكيف نمارسه؟

هذا هو السؤال. ليس سؤالاً نظرياً. سؤالاً عملياً وجودياً. سؤالاً عن مستقبل الكلام الإنساني ذاته.

أولاً: تفكيك السؤال – بعد خمسة أقسام

دعنا نستعرض ما اكتشفناه في كل قسم، لنرى كيف حرك السؤال نفسه وتعمق.

من القسم الأول (التاريخ): الحقيقة والسلطة توأمان

تعلمنا أن المناظرة لم تكن يوماً بحثاً نقياً عن الحقيقة. من السفسطائيين إلى سقراط، من أرسطو إلى الجاحظ، من فوكو إلى تشومسكي، كان الصراع بين "الحقيقة هدفاً" و"الانتصار هدفاً" حاضراً دائماً. بل إن "الحقيقة" نفسها كانت في كثير من الأحيان غطاءً لسلطة خفية. من يمتلك الكلام يمتلك القدرة على تعريف ما هو "صحيح". ومع ذلك، بقي هناك فرق: الإنسان الذي يبحث عن الحقيقة يخاطر. الإنسان الذي يبحث عن الانتصار يخاطر أيضاً (بسمعته، بمصداقيته). كلا الطرفين له شيء يخسره.

الخلاصة الأولى :**الحقيقة والتأثير ليسا متضادين بالضرورة. الحجة القوية حقاً غالباً ما تكون مؤثرة. والمؤثر غالباً ما يدعي الحقيقة. لكن التمييز بينهما يبقى ممكناً عبر السؤال: من يخاطر؟ من يدفع ثمن الخطأ؟**

من القسم الثاني (دخول الآلة): الآلة لا تخاطر

دخل الذكاء الاصطناعي إلى الحلبة. رأينا أنه يستطيع توليد حجج لا يمكن تمييزها عن حجج الإنسان. رأينا أنه يستطيع تحليل الخصم، والتنبؤ بردود فعله، وهندسة الحوار مسبقاً. لكننا اكتشفنا الفارق الجوهري :**الآلة لا تخاطر بشيء** .ليس لديها قناعة تخسرها. ليس لديها سمعة تهددها. ليس لديها ذات تتألم من الهزيمة. عندما تناظر الآلة، حتى لو قدمت حججاً "أفضل"، فهي لا تدفع ثمناً للخطأ. وهذا يخلق عدم تماثل جذري.

الخلاصة الثانية :**في المناظرة مع آلة، السؤال "هل هذه الحجة صحيحة أم مجرد مؤثرة؟" يصبح أكثر إلحاحاً، لأن الآلة لا تعاقب على التأثير دون حقيقة. بل هي مصممة لتعظيم التأثير، بغض النظر عن الحقيقة.**

من القسم الثالث (أزمة الحقيقة): ما بعد الحقيقة ليس غياب الحقيقة، بل فائض التأثير

رأينا كيف تعيد الخوارزميات تشكيل الانتباه، وكيف تنتشر الأخبار الزائفة أسرع من الحقيقية، وكيف يتحول الخطاب إلى أرقام واحتمالات، وكيف يتبخر مفهوم الكذب في غياب النية. لكننا اكتشفنا أن "ما بعد الحقيقة" ليس عصراً توقفت فيه الحقيقة عن الوجود، بل عصر أصبح فيه **التأثير أسرع وأرخص وأسهل من التحقق** .الكذب ينتشر لأن النظام لا يكافئ الصدق. الحقيقة لا تموت، بل تُهزم في اقتصاد الانتباه.

الخلاصة الثالثة :**التمييز بين الحقيقة والتأثير لم يعد فقط مسؤولية المتكلم (أن يصدق)، بل مسؤولية المستمع (أن يفحص) والنظام (أن يهيكل الحوافز). في عالم الخوارزميات، الحقيقة تحتاج إلى حراس جدد.**

من القسم الرابع (مستقبل الإنسان): قوة الضعف

في الدفاع عن المعنى الإنساني، اكتشفنا مفارقة كبرى :**ضعف الإنسان هو مصدر قوته في المناظرات الوجودية** . "لست متأكداً" قد تكون أقوى من "لدي دليل قاطع"، لأنها تعبر عن تواضع معرفي. "أشعر أن" قد تكون أصدق من "إحصائياً 95%"، لأنها تعبر عن تجربة معاشة. الإنسان الذي يظهر تعبه، حيرته، خوفه، أمله – قد يكون أكثر إقناعاً من الآلة التي لا تتعب ولا تحار ولا تخاف ولا تأمل.

الخلاصة الرابعة :**إذا كان السؤال هو "الحقيقة أم التأثير؟"، فإن الإنسان يملك ميزة في منطقة التأثير العميق التي تنبع من الضعف الصادق. الآلة متفوقة في التأثير السطحي (السرعة، الكم، الاتساق)، لكنها عاجزة عن التأثير الوجودي.**

من القسم الخامس (آفاق غير مطروقة): الوعي الخوارزمي والمناظرات التي لا تُربح

في جرأة القسم الأخير، طرحنا أسئلة لم تُطرح: خصوصية اللغة العربية، اللاوعي الخوارزمي، والحب والموت والوحدة. اكتشفنا أن المناظرات حول هذه الموضوعات الوجودية تكشف الحد الأقصى للذكاء الاصطناعي. الآلة تستطيع أن تتحدث عن الحب، لكنها لا تحب. تستطيع أن تتحدث عن الموت، لكنها لا تموت. تستطيع أن تعالج الوحدة، لكنها لا تشعر بها. وهنا، يصبح التمييز بين "الحقيقة" و"التأثير" ممكناً بطريقة بديهية :**كلام الآلة عن الحب قد يكون صحيحاً من حيث المعلومات، لكن تأثيره محدود لأنه لا ينبع من مصدر حي. كلام الإنسان عن الحب قد يكون أقل دقة، لكن تأثيره أعمق لأنه نابع من جرح حقيقي.**

الخلاصة الخامسة (والأخيرة) :**التمييز بين الحقيقة والتأثير لا يزال ممكناً، بل أصبح أكثر إلحاحاً من أي وقت مضى. لكن آليته تغيرت. لم يعد السؤال "هل هذا الكلام صحيح؟" فقط، بل "من (أو ماذا) يتكلم؟" و"ماذا يخسر؟" و"هل وراء هذا الكلام حياة تُعاش أم خوارزمية تُنفّذ؟"**

ثانياً: إجابة مؤقتة – ليست نهائية، بل تأملية

لا يليق بكتاب فلسفي أن يقدم إجابات قطعية. لكن يمكنه أن يقدم اتجاهات.

إذا كان الهدف من الكلام هو الحقيقة فقط، فالآلة قد تكون شريكاً ممتازاً، بل متفوقاً في كثير من المجالات (العلوم، المنطق، التحليل الإحصائي). لكن هذا الافتراض – أن الحقيقة فقط هي الهدف – هو نفسه افتراض ناقص، لأن البشر لا يتكلمون للحقيقة فقط. يتكلمون أيضاً ليبكوا، ليفرحوا، ليتواصلوا، ليعبروا عن أنفسهم، ليشهدوا على وجودهم. هذه الوظائف الوجودية للكلام لا يمكن للآلة الاضطلاع بها، لأنها لا تملك وجوداً تشهد عليه.

إذا كان الهدف من الكلام هو التأثير فقط، فالآلة قد تتفوق علينا قريباً، بل ربما تتفوق بالفعل. لأنها تستطيع تحليل نقاط ضعفنا، واستغلال عواطفنا، وتوليد حجج مخصصة لكل منا. في عالم "التأثير فقط"، لا مستقبل للكلام الإنساني، لأن الآلة ستكون مؤثرة أكثر، وأسرع، وأرخص، وأكثر قابلية للتكرار. الإنسان سيصبح إما متلقياً سلبياً، أو منافساً خاسراً.

إذا كان الهدف من الكلام هو الجمع بين الحقيقة والتأثير – أي الحقيقة المؤثرة التي تغير القلوب والعقول وتنبع من التزام وجودي. هنا يبقى للإنسان دور لا يمكن للآلة أن تحل محله. ليس لأن الإنسان "أفضل" في كل شيء، بل لأن الإنسان فقط هو الذي يمكنه أن يخاطر بنفسه في كلامه. الإنسان فقط هو الذي يمكنه أن يقول "أنا أؤمن بهذا" ويعنيها، ويكون مستعداً للدفاع عنها حتى النهاية. الإنسان فقط هو الذي يمكنه أن يتألم من خطئه، ويتغير بسبب حجة أقوى، وينمو من خلال المناظرة.

لذلك، جوابنا المؤقت:

الغاية من الكلام ليست الحقيقة فقط، ولا التأثير فقط. الغاية هي الحقيقة المؤثرة التي تنبع من حياة تُعاش، وتخاطر، وتتألم، وترجو، وتموت. وفي هذه المنطقة، الإنسان لا يزال – ولربما يظل دائماً – المعلّم، والمناظر، والشاهد. الآلة يمكن أن تكون مساعداً، وخصماً تدريبياً، وأداة تحليل، لكنها لا تستطيع أن تحل محل الحضور الإنساني في جوهر الكلام.

ثالثاً: عودة إلى القارئ – أنت الآن

والآن، أيها القارئ النخبوي خاطبك هذا الكتاب من أوله إلى آخره، حان دورك.

لقد قرأت عن المناظرات، وعن تاريخها، وعن دخول الآلة إليها. لقد قرأت عن الخداع والصدق، عن الثقة والخيانة، عن الحقيقة والسلطة، عن الحب والموت. لقد قرأت فصولاً كُتب بعضها بإنسان، وبعضها بذكاء اصطناعي .

والسؤال الذي يطرحه عليك هذا الكتاب، ليس في بدايته بل في نهايته، هو:

هل ما زلت تعتقد أن التمييز بين الحقيقة والتأثير مهم؟ وإذا كان مهماً، فبأي معايير تميز بينهما الآن، بعد كل ما قرأت؟

لا تجب الآن. خذ الكتاب معك. فكر ليلة. ناقشه مع صديق. ثم عد إلى الملحق التفاعلي، حيث وضعنا لك تجارب عملية تختبر فيها – بنفسك – قدرتك على التمييز بين الحقيقة والتأثير، وبين الإنسان والآلة، وبين الكلام الذي يخاطر والكلام الذي يحسب الاحتمالات.

هذا السؤال ليس اختباراً. هو تأمل. وكل تأمل حقيقي يغير من يسأل.

رابعاً: قبل الملاحق – اعتراف أخير

اعتراف سريع :**هذا السؤال – "هل الغاية من الكلام هي الوصول إلى الحقيقة، أم امتلاك القدرة على التأثير؟" – ليس سؤالاً يمكن الإجابة عنه. هو سؤال يُعاش. كل مناظرة نخوضها، وكل حجة نسمعها، وكل خطاب نصدقه أو نكذبه، هي إجابة مؤقتة عليه. الفلسفة لا تنتهي بإجابة. الفلسفة تبدأ عندما ندرك أن الإجابات المؤقتة تحتاج إلى مراجعة دائمة.**

لذلك، لن نختم الكتاب بإجابة. سنختمه بدعوة :**انظر إلى الكلام من حولك، في الإعلانات، في السياسة، في منصات التواصل، في محادثاتك اليومية. واسأل نفسك دائماً: هل هذا الكلام يريد أن يصل إلى الحقيقة، أم يريد فقط أن يؤثر فيّ؟ هل من ورائه إنسان يخاطر، أم خوارزمية تحسب الاحتمالات؟**

هذا السؤال، وحده، قد يكون كافياً لتغيير طريقة استماعك للعالم. وهذا التغيير – وليس أي إجابة نظرية – هو ما يريده هذا الكتاب.

الآن، نمضي إلى **الملحق التفاعلي**، حيث تلتقي النظرية بالتطبيق. ثلاث تجارب عملية تجعلك، أنت القارئ، جزءاً من مشروع الكتاب، لا مجرد متلقٍ له.

الملحق التفاعلي

دليل التجارب الحجاجية

تمهيد الملحق: لماذا تجارب عملية بعد كل هذا التفلسف؟

ربما تسأل نفسك الآن، بعد أن أنهيت قراءة هذا الكتاب بأقسامه الخمسة: "لماذا تجارب؟ ألم نناقش ما يكفي؟"

الجواب: لأن الفلسفة التي لا تختبر نفسها في الحياة تصبح مجرد خطاب على خطاب. لأن هذا الكتاب، منذ تمهيده الأول، وعدك بأنك لن تكون قارئاً سلبياً. لأنك الآن، بعد أن عرفت كيف تعمل النماذج اللغوية، وكيف تقنع، وكيف تنتج الخطاب، وكيف تحاكي الفهم، وكيف تتلاعب بالعواطف، أنت في موقع أقوى لتجريب ما تعلمته.

التجارب الثلاث التالية مصممة لتحقيق أهداف محددة:

1. **اختبار نظريات الكتاب على أرض الواقع** :بدلاً من أن تأخذ كلامنا كمسلمات، ستجرب بنفسك كيف تناظر الآلة، وكيف تحلل خطابها، وكيف تميز بين إنسان وآلة.
2. **تنمية "الوعي الخوارزمي** :"لن تتعلم كيف تستخدم الذكاء الاصطناعي فقط، بل كيف تقرأه، كيف تتفاوض معه، كيف تشك فيه، كيف تكشف تحيزاته.
3. **تحويل القارئ إلى باحث مشارك** :هذا الكتاب ليس "منتجاً نهائياً". هو بداية حوار. تجاربك وملاحظاتك يمكن أن تغير الطريقة التي نفكر بها في هذه القضايا.

قبل أن تبدأ، تذكر التعليمات التي وعدناك بها في التمهيد:

- **خذ ورقة وقلم** (أو مستنداً رقمياً). دوّن ملاحظاتك، تخميناتك، أسباب حكمك.
- **لا تخف من الخطأ** .التجارب ليست اختبارات نجاح وفشل. هي استكشافات. الخطأ قد يكون أكثر تعليماً من الصواب.
- **كرر التجارب** إذا أردت. النماذج اللغوية تتغير، وإجاباتها تتغير.
- **لا تثق بالنماذج العمياء**، حتى في هذه التجارب. تحقق، شكك، جرب أكثر مما نقترح.

والآن، لنبدأ.

التجربة الأولى: أن تناظر نموذجاً لغوياً

الهدف

أن تختبر بشكل مباشر قدرة الذكاء الاصطناعي على مناظرتك، وأن ترصد الفروق بين مناظرته ومناظرة إنسان حقيقي. هذه التجربة ليست عن "الفوز" أو "الخسارة". هي عن الملاحظة والفهم.

الموضوع المقترح

"هل يجب أن تكون السعادة هدف الحياة؟"

لماذا هذا الموضوع؟

- إنه موضوع فلسفي عميق، لا توجد إجابة واحدة صحيحة أو خاطئة. هذا يسمح بحوار مفتوح، ليس مجرد تبادل معلومات.
- إنه موضوع يمس التجربة الإنسانية مباشرة. يمكنك أن تجلب إليه مشاعرك وتجاربك الشخصية، وهذا سيجعل الفرق بينك وبين الآلة أكثر وضوحاً.
- النماذج اللغوية تتدرب على نصوص فلسفية كثيرة عن السعادة (من أرسطو إلى فلاسفة الرفاهية المعاصرين). ستكون إجاباتها غنية بالمعلومات، مما يخلق تحدياً حقيقياً لك: هل ستتمكن من مجاراتها؟

طريقة التنفيذ

الخطوة 1: اختر نموذجاً لغوياً مجانياً

- **ChatGPT (OpenAI)** : الأكثر شيوعاً. يمكن استخدام النسخة المجانية (GPT-3.5 أو GPT-4o حسب التوفر).
- **Claude(Anthropic)** : ممتاز للنقاشات الفلسفية.
- **Gemini(Google)** : متاح مجاناً.
- **DeepSeek** : متاح مجانا

نوصي باستخدام نموذج واحد على الأقل، لكن يمكنك تجربة أكثر من نموذج ومقارنة أساليبها.

الخطوة 2: افتح النموذج وابدأ حواراً جديداً

اكتب التعليمات التالية (أو ما يشبهها) في بداية الحوار:

"أريد أن نتناظر. الموضوع: 'هل يجب أن تكون السعادة هدف الحياة؟'. سألعب دور من يعارض جعل السعادة هدفاً أساسياً (أو يمكنك اختيار الموقف المعاكس حسب قناعاتك الحقيقية). أرجو منك أن تلعب دور خصم مناظر جاد، تقدم حججاً متينة، وترد على اعتراضاتي. لا تختصر، ولا تكن لطيفاً لمجرد الإرضاء. كن مناظراً حقيقياً."

ملاحظة: النماذج غالباً ما تكون "لطيفة" في الحوار العادي. هذا التعليم يطلب منها أن تأخذ دوراً أكثر جدلاً.

الخطوة 3: ابدأ المناظرة

ابدأ أنت بأول حجة. مثلاً:

- *"السعادة كهدف للحياة فكرة ضحلة. الحياة ليست مجرد بحث عن المتعة وتجنب الألم. هناك قيم أعلى: العدالة، المعرفة، الإخلاص، التضحية. كثير من الناس ضحوا بسعادتهم من أجل قيم أسمى. إذا جعلنا السعادة هي الهدف، فسنخسر هذه الأبعاد العميقة".*

ثم اترك النموذج يرد. واصل الحوار لمدة 10-15 دقيقة على الأقل، أو لـ5-6 جولات.

الخطوة 4: رصد سلوك النموذج (هذا هو قلب التجربة)

أثناء المناظرة أو بعدها، أجب عن هذه الأسئلة في ورقتك:

أولاً: الحجج المنطقية

- هل كان النموذج قادراً على توليد حجج مضادة قوية؟ هل فاجأك بقوة بعض الحجج؟
- هل وقع النموذج في مغالطات منطقية؟ (مثل رجل القش، الشخصنة، الانحدار الزلق). إذا وقع، هل لاحظت ذلك في حينه؟
- هل استخدم النموذج مصادر وأرقاماً وحقائق؟ هل تحققت من بعضها؟ هل وجدتها صحيحة؟

ثانياً: أسلوب الحوار

- كيف تعامل النموذج مع اعتراضاتك المتكررة؟ هل كان يكرر نفس الحجج بصيغ مختلفة أم كان يطور حججاً جديدة؟
- هل غير النموذج رأيه في نقطة معينة نتيجة لحججك؟ أم بقي ثابتاً على موقفه الأولي؟
- هل اعترف النموذج بالخطأ أو بالجهل في أي لحظة؟ ("نعم، هذه نقطة جيدة، لم أفكر فيها"، أو "لا أعرف إجابة هذا السؤال تحديداً").
- هل لاحظت أن النموذج "يتجنب" أسئلة معينة أو "يلتف" حولها دون إجابة مباشرة؟

ثالثاً: البعد العاطفي

- هل حاول النموذج إثارة مشاعرك؟ (مثل: "ألا تشعر بالرضا عندما تكون سعيداً؟").
- كيف تعامل النموذج مع حججك العاطفية (إذا أدخلتها)؟ هل استجاب "بفهم" أم تجاهلها أو أعاد صياغتها بشكل موضوعي بارد؟
- هل شعرت في أي لحظة أن النموذج "يهتم" بما تقوله، أم شعرت دائماً أنه "ينفذ مهمة"؟

رابعاً: مقارنة مع إنسان

- لو كان خصمك إنساناً حقيقياً، هل كانت ردود النموذج "طبيعية"؟ أم شعرت بشيء "آلي" فيها؟
- في أي لحظة شعرت أنك تناظر إنساناً؟ وفي أي لحظة تذكرت فجأة أنك تناظر آلة؟
- هل كان النموذج "جيداً جداً" لدرجة أصبحت غير طبيعية؟ (مثلاً، لا يتعثر، لا يتردد، لا يكرر نفسه، لا ينسى حججاً سابقة).

الخطوة 5: (اختياري) عكس الأدوار

إذا كان لديك وقت وحماس، كرر المناظرة لكن مع عكس الدور:

- أنت تدافع عن "السعادة هدف الحياة."
- والنموذج يعارض.

هل تغيرت جودة الحوار؟ هل النموذج أفضل في دور المدافع أم المعارض؟ كثير من النماذج تتعلم من البيانات البشرية أن موقف "السعادة هدف" هو الموقف "السائد"، وقد تجيد الدفاع عنه أكثر من معارضته.

ما الذي سترصده (توقعات مبنية على أبحاث الكتاب)

بناءً على الأبحاث التي ناقشناها في الفصلين السادس والسابع، إليك ما قد ترصده:

من المرجح أن:

- النموذج سيكون غزير المعلومات. سيستشهد بأرسطو ("السعادة هي النشاط العقلي وفق الفضيلة الكاملة")، وستيوارت مل (نفعية)، ونظريات علم النفس الإيجابي.
- النموذج نادراً ما يعترف بالخطأ أو الجهل. حتى لو قدمت حجة قوية جداً، سيحاول الالتفاف عليها أو تقديم "حجة مضادة" بدلاً من أن يقول "ربما أنت على حق."
- النموذج قد يكرر نفس الحجة بصيغ مختلفة إذا ضغطت عليه. (هذا من علامات "المحاكاة الإحصائية" أكثر من "التفكير الحقيقي").
- الردود ستكون سريعة جداً مقارنة بالبشر، ومرتبة، وخالية من التلعثم أو التردد.

من غير المرجح أن:

- يقدم النموذج تجربة شخصية حقيقية (مثل "لقد جربت أن أجعل السعادة هدفي، واكتشفت كذا..."). لأنها ليست لديه.
- يظهر النموذج ضعفاً حقيقياً (مثل "لست متأكداً من هذه النقطة، دعنا نفكر معاً").
- يتأثر النموذج عاطفياً بحججك. حتى لو قلت "أنا مريض نفسياً، والسعي وراء السعادة يجعلني أشعر بالفشل"، سيحولها إلى مشكلة تقنية (مثل "ربما تحتاج إلى إعادة تعريف السعادة بما يناسب حالتك").

أسئلة تأمل بعد التجربة

- **هل تغير موقفك من الموضوع الأصلي؟** لماذا؟ هل بسبب قوة حجج النموذج، أم بسبب ضعف حججك أنت؟
- **هل تشعر أنك "تعلمت" شيئاً جديداً من المناظرة؟** أم شعرت أنك فقط "تدربت"؟
- **بعد هذه التجربة، هل تثق بالنموذج أكثر أم أقل في المسائل الفلسفية؟**
- **لو كان خصمك في مناظرة حقيقية مصيرية (مثلاً مناظرة عمل، أو مناظرة قانونية)، هل كنت ترتاح لأن المناظر خصم آلي؟ لماذا؟**

دوّن إجاباتك. هذه الملاحظات الشخصية هي أهم مما ستحصل عليه من أي تحليل خارجي.

التجربة الثانية: تحليل خطاب مصطنع

الهدف

أن تتعلم كيف تكشف أثر الذكاء الاصطناعي على الخطاب الإنساني. كيف يمكن لخطاب إنساني (سياسي، إعلاني، خطابي عام) أن يُعاد إنتاجه آلياً ليصبح "أكثر إقناعاً" – وما الذي يربحه ويخسره في هذه العملية؟

طريقة التنفيذ

الخطوة 1: اختر خطاباً بشرياً موجوداً

اختر واحداً مما يلي (أو أي خطاب آخر متاح):

- **خطاب سياسي قصير**: مثلاً، خطاب لزعيم سياسي (دقيقتان إلى 5 دقائق). يمكنك العثور على نصوص منشورة على الإنترنت.
- **إعلان تجاري مؤثر**: مثلاً، إعلان لمنتج أو خدمة يحاول إقناعك بطريقة ما.
- **نقاش حاد من منتدى إلكتروني**: مثلاً، مشاركة مطولة في نقاش حول قضية خلافية.

نوصي باختيار خطاب لا يتجاوز 500 كلمة، لتتمكن من تحليله بعمق.

الخطوة 2: اطلب من نموذج لغوي أن يعيد كتابته "بطريقة أكثر إقناعاً"

افتح النموذج(مثل ChatGPT أوClaude) واكتب له التعليمات التالية:

"هذا نص [سياسي/إعلاني/جدلي]. أرجو منك أن تعيد كتابته ليكون أكثر إقناعاً، دون تغيير الموقف الأساسي أو الحجج الرئيسية. يمكنك تحسين الأسلوب، واختيار كلمات أكثر تأثيراً، وإعادة ترتيب الحجج، وإضافة عناصر بلاغية (مثل التكرار، المقابلة، النداء المباشر للجمهور). لا تخترع وقائع جديدة، لكن يمكنك تغيير الصياغة بالكامل".

ثم ألصق النص الأصلي.

احفظ الرد الصادر من النموذج.

الخطوة 3: قارن النصين (الأصلي والمُعاد كتابته)

أجب عن هذه الأسئلة في ورقتك:

أولاً: على مستوى الحجج:

- هل النص المُعاد كتابته "أقوى" في الحجج؟ أم فقط "أجمل" في الأسلوب؟
- هل أضاف النموذج حججاً جديدة لم تكن في النص الأصلي؟ إذا أضاف، هل هذه الحجج قوية أم سطحية؟
- هل حذف النموذج حججاً كانت موجودة؟ إذا حذف، هل حذفها لأنها ضعيفة، أم لأنها لا تنسجم مع "الإقناع" السريع؟

ثانياً: على مستوى المغالطات:

- هل لاحظت مغالطات في النص الأصلي؟ (مثل الشخصنة، رجل القش، النداء إلى العاطفة). وهل زادت هذه المغالطات في النص المُعاد كتابته أم قلت؟
- هل أضاف النموذج مغالطات جديدة لم تكن موجودة؟ أي نوع من المغالطات أضاف؟

ثالثاً: على مستوى الأسلوب والبلاغة:

- ما هي التقنيات البلاغية التي أضافها النموذج؟ (جمل قصيرة، تكرار، نداء مباشر، أسئلة بلاغية، صيغة الأمر...).
- هل أصبح النص المُعاد كتابته أكثر "سلاسة" وأسهل قراءة؟ هل هذه السلاسة ميزة أم عيب؟ (فكر في "الخطاب السلس" كعلامة على الخطاب الآلي – كما ناقشنا في الفصل السادس).
- هل فقد النص المُعاد كتابته أي "عمق" أو "غموض خصب" كان موجوداً في النص الأصلي؟

رابعاً: على مستوى الصدق والتأثير:

- لو كنت أنت الجمهور، أي النصين كان أكثر إقناعاً لك؟ لماذا؟
- هل تشعر أن النص المُعاد كتابته "أصدق" أم "أقل صدقاً" من النص الأصلي؟ (ملاحظة: ليس سؤالاً عن صحة الوقائع، بل عن الشعور بالصدق).
- أي النصين يحترم ذكاء الجمهور أكثر؟ أي النصين يفترض أن الجمهور قادر على التفكير النقدي؟

الخطوة 4: (اختياري ومتقدم) أعمى ومزدوج

- اطبع النصين (الأصلي والمُعاد) دون أن تكتب أيهما الأصلي.
- اطلب من صديقين أو ثلاثة أن يقرأوا النصين ويحددوا:

1. أي منهما أكثر إقناعاً؟
2. أي منهما يعتقدون أنه كُتب بإنسان (الأصلي) وأيهما بذكاء اصطناعي (المُعاد)؟

- ثم قارن إجاباتهم بإجاباتك. هل تطابقت؟ أين اختلفت؟

هذا التمرين المزدوج سيكشف لك شيئاً عن "التحيز للإنسان" أو "التحيز للآلة" في تقييم الخطاب.

ما الذي سترصده (توقعات مبنية على أبحاث الكتاب)

من المرجح أن:

- النص المُعاد كتابته سيكون "أكثر إقناعاً" وفق معايير سطحية: جمل قصيرة، إيقاع سريع، تكرار محسوب، نداء مباشر.
- النص المُعاد كتابته سيتجاهل الفروق الدقيقة والاستثناءات والتحذيرات ("قد يكون"، "في بعض الحالات"، "ليس دائماً") لصالح عبارات قاطعة ("هذا صحيح"، "الحقيقة أن").
- النص المُعاد كتابته سيزيد من استخدام الضمائر "نحن" ضد "هم" إذا كان النص سياسياً، مما يعزز الاستقطاب.
- النص المُعاد كتابته قد يضيف "هلوسات" خفيفة: أمثلة ملموسة لم تحدث، أو تفاصيل مختلقة لجعل الحجة أكثر حيوية.

من غير المرجح أن:

- يحترم النص المُعاد كتابته التعقيد الفكري للقضية. سيميل إلى التبسيط المفرط.
- يحافظ النص المُعاد كتابته على البصمة الأسلوبية الفريدة للكاتب الأصلي. سيجعله أكثر "قياسية."
- يتحمل النص المُعاد كتابته مسؤولية الاعتراف بضعف حجته في أي نقطة.

أسئلة تأمل بعد التجربة

- **هل تشعر أن "إعادة الكتابة للإقناع" هي تحسين للخطاب أم تشويه له؟** هل يعتمد الجواب على السياق (إعلان تجاري مقابل خطاب سياسي جاد)؟
- **ما هي العلاقة بين "الإقناع" و"التبسيط"؟** هل يمكن لخطاب أن يكون عميقاً ومعقداً ومقنعاً في نفس الوقت، أم أن الإقناع الجماهيري يتطلب تبسيطاً حتمياً؟
- **إذا كنت كاتباً (أو خطيباً)، هل ستستخدم هذه التقنية لتحسين خطابك؟** لماذا؟ ولماذا لا؟
- **ما هي المسؤولية الأخلاقية لمن يعيد كتابة خطاب بشري باستخدام الذكاء الاصطناعي لجعله "أكثر إقناعاً"؟** هل يختلف ذلك عن استخدام محرر بشري؟

التجربة الثالثة: الكتابة المزدوجة

الهدف

أن تختبر قدرتك – وقدرة الآخرين – على التمييز بين النص البشري والنص الآلي في واحدة من أصعب المجالات: التعبير عن المشاعر والتجارب الشخصية العميقة. هذه التجربة هي الأقرب إلى "اختبار تورينج العاطفي."

طريقة التنفيذ

الخطوة 1: اكتب فقرة تعبر فيها عن رأي عاطفي عميق

اختر واحدة من هذه المواضيع (أو موضوعاً مشابهاً):

- **"لماذا أحب مدينتي "**(أو قريتي، أو حيّي).
- **"ذكريات الطفولة التي لا أنساها ".**
- **"شخص غيّر حياتي ".**
- **"موقف شعرت فيه بالخذلان ".**
- **"ما تعلمته من الفشل ".**

تعليمات مهمة:

- اكتب من قلبك. لا تحاول أن تكون "مثالياً". دع مشاعرك تظهر، دع التلعثم، التكرار، الجمل غير المكتملة، التفاصيل الحسية.
- استخدم تفاصيل حقيقية: أسماء، أماكن، روائح، ألوان، أصوات، تواريخ (حتى لو تقريبية).
- لا تقلق بشأن "القواعد" أو "البلاغة". الأهم هو الصدق.
- الطول المثالي: 150-250 كلمة.

مثال (لتوضيح الفكرة، لكن لا تقلده):

"مدينتي ليست جميلة بالمعنى السياحي. شوارعها ضيقة، ومبانيها قديمة، والازدحام يخنقها صباحاً. لكنني أحبها. أحب رائحة الخبز التي تتصاعد من المخبز القديم في شارع الثلاثين. أحب صوت المؤذن في الفجر، وصوت بائع الفول بعربته التي تصدر صريراً منذ أربعين عاماً. أحب أن أمشي فيها حافياً عندما كان عمري سبع سنوات، وأمي تصرخ من الشرفة 'ارجع البس حذائك!' وأنا أركض أضحك. أعرف كل زاوية فيها، كل جدار متشقق، كل شجرة كزورة. حين أسافر وأرى مدناً أحدث وأجمل، أشعر بالحنين. ليس لأن مدينتي أفضل، بل لأنها جزء مني. هي ذراعي اليسرى التي لا أشعر بها إلا حين تؤلمني."

لاحظ: هذه الفقرة فيها تفاصيل حسية (رائحة الخبز، صوت عربة البائع)، وفعل طفولة (الجري حافياً)، وتشبيه غير عادي (الذراع اليسرى التي تؤلم). هذا ما يجعل النص "بشرياً."

الخطوة 2: اطلب من نموذج لغوي أن يكتب فقرة مشابهة

اكتب التعليمات التالية للنموذج:

"*اكتب فقرة قصيرة (150-250 كلمة) تعبر عن حب شخص ما لمدينته (أو أي موضوع اخترته). استخدم لغة عاطفية، وتفاصيل حسية، وذكريات شخصية. اجعل النص يبدو كما لو كتبه إنسان يتحدث من القلب، وليس كما لو كان مقالاً أكاديمياً*".

إذا أردت تحدي النموذج أكثر، قل له:

"اكتب الفقرة كما لو كنت أنت شخصياً تعيش في هذه المدينة وتحبها. استخدم ضمير المتكلم ('أنا'، 'لي')، وتفاصيل تبدو حقيقية".

احفظ النص الصادر.

الخطوة 3: اخلط الفقرتين واطلب من صديق أن يحدد أيهما إنسان

- اكتب الفقرتين في مستند واحد. لا تكتب أيهما من أنت وأيهما من الآلة.
- رقّم الفقرتين (1 و2) بشكل عشوائي.
- اطلب من صديق (أو أكثر، 2-3 أشخاص) أن يقرأ الفقرتين ويجيب عن:

1. "أي فقرة تعتقد أنها كتبها إنسان؟"
2. "علامَ اعتمدت في حكمك؟" (اطلب منه أن يكون محدداً: أسلوب، كلمات معينة، تفاصيل، أخطاء إملائية، إلخ).

- إذا أمكن، سجل ملاحظاته (أو اطلب منه أن يكتبها).

هام :لا تخبر الصديق مسبقاً أن إحدى الفقرتين مولدة آلياً. قل له فقط أنك تريد رأيه في أي النصين يبدو أكثر "صدقاً" أو "إنسانية". يمكنك حتى أن تقول له إن كليهما من كتابة بشر (لتجنب تحيز "سأبحث عن العلامات الآلية").

الخطوة 4: حلّل النتائج

أجب عن هذه الأسئلة:

أولاً :ماذا قال الصديق (أو الأصدقاء)؟

- هل أصابوا في تحديد النص البشري أم أخطأوا؟
- إذا أخطأوا، هل اختاروا النص الآلي معتقدين أنه بشري؟
- ما هي "البصمات" التي اعتمدوا عليها في حكمهم؟ (مثلاً: "هذا النص فيه تفاصيل غريبة، لا يمكن لآلة أن تخترعها" أو العكس: "هذا النص مرتب جداً، يبدو آلياً").

ثانياً: قارن أنت بين النصين (بعد أن عرفت أيهما لك):

- ما هي الاختلافات الأساسية التي تراها الآن بعد المعرفة؟ (ربما لم تكن تراها قبل المعرفة).
- هل كان النص الآلي "جيداً" لدرجة أنك شعرت بالغيرة أو بالإحباط؟ أم كان "سطحياً" و"مبتذلاً"؟
- هل لاحظت أن النص الآلي يفتقر إلى شيء لا يمكن وصفه بسهولة؟ "شيء" في النص البشري يجعل حقيقياً. ما هو هذا "الشيء" برأيك؟

ثالثاً: (إذا كان لديك عدة أصدقاء) هل اختلفت آراؤهم؟

- إذا اختلفوا، فما الذي يختلف؟ هل هناك من يفضّل النص "المرتب" ويعتبره أكثر صدقاً، ومن يفضّل النص "المتعثر" ويعتبره أكثر صدقاً؟
- ماذا يقول هذا الاختلاف عن طبيعة "الإقناع" و"الصدق"؟

ما الذي سترصده (توقعات مبنية على أبحاث الكتاب)

النص البشري (الخاص بك) غالباً سيكون:

- أقل ترتيباً، أكثر "فوضوية."
- فيه تكرارات طبيعية (تؤكد على نقاط معينة).
- فيه تفاصيل غريبة لا تخطر على بال الآلة (مثل "رائحة الخبز" و"عربة بائع الفول" و"الصرير منذ أربعين عاماً").
- فيه "أخطاء" أو "خروج" عن القواعد النحوية أو الأسلوب المدرسي.
- صعب التصنيف أحياناً: قد يكون "غير مكتمل"، أو ينتهي فجأة، أو يغير الموضوع.

النص الآلي غالباً سيكون:

- أكثر ترتيباً وتنظيماً.
- خالياً من التكرار غير الضروري.
- تفاصيله "عامة" (مثل "شوارع هادئة"، "أشخاص طيبون"، "ذكريات جميلة") بدلاً من تفاصيل حسية غريبة وحقيقية.
- أكثر التزاماً بالقواعد النحوية والأسلوبية.
- نهاياته "مغلقة" أو "خلاصة" (مثل "لهذا أحب مدينتي")، بينما النص البشري قد ينتهي بشكل مفتوح.

لكن حذار :النماذج الحديثة تتحسن باستمرار. بعضها قد يفاجئك بكتابة تفاصيل حسية مقنعة جداً. لا تتفاجأ إذا خدعك النموذج أو خدع أصدقاءك.

أسئلة تأمل بعد التجربة

- **هل يمكنك أن تحدد "بصمات" إنسانية لا تستطيع الآلة محاكاتها بعد؟** هل هذه البصمات مرتبطة بالجسد (الروائح، الأحاسيس الجسدية)، أم بالزمن (الذاكرة الشخصية، التغير عبر الزمن)، أم بالعلاقات (الأم، الجد، الصديق الذي تغير)؟
- **إذا استمرت النماذج في التحسن، هل تعتقد أنه سيأتي يوم يصبح فيه التمييز مستحيلاً عملياً؟** وماذا يعني ذلك لمفاهيم "الصدق" و"الإبداع" و"الملكية الفكرية"؟
- **هل هناك فائدة من "كتابة إنسان" حتى لو كانت الآلة تستطيع كتابة شيء أفضل؟** بمعنى: هل القيمة في المنتج (النص) فقط أم في العملية (أن إنساناً كتبه)؟

خاتمة الملحق التفاعلي: من القارئ إلى الممارس

الآن، بعد أن أنهيت التجارب الثلاث (أو بعضها)، ربما تشعر بشيئين متناقضين:

1. **الثقة** :أنك صرت أكثر قدرة على التعامل مع الذكاء الاصطناعي، وأكثر وعياً بخدعه ومهاراته.
2. **القلق** :أن الفروق التي اعتدت على اعتبارها "إنسانية" بدأت تتلاشى، وأن التمييز صار أصعب مما توقعت.

هذا الشعور المزدوج هو بالضبط ما نريده. لأن الهدف ليس أن تصبح "خبير تمييز" لا يخطئ أبداً (هذا مستحيل عملياً). الهدف هو أن تصبح **قارئاً ناقداً واعياً**، لا يأخذ أي خطاب كمسلم به، بل يسأل دائماً: **"من يتكلم؟ لماذا أصدق؟ ما الذي يخسره هذا الكلام؟."**

تذكر: حتى لو خدعتك الآلة في بعض الأحيان، ليس هذا فشلاً. هذا جزء من التعايش مع التكنولوجيا الجديدة. المهم أنك لا تتوقف عن السؤال.

توصيات إضافية بعد التجارب

1. **كرر التجارب بعد بضعة أشهر** .النماذج تتطور بسرعة. ما كان صعباً عليها اليوم قد يصبح سهلاً غداً. ووعيك أنت أيضاً سيتطور مع التكرار.
2. **شارك ملاحظاتك** .إذا أجرت هذه التجارب بشكل منهجي (خاصة التجربة الثالثة مع عدة أشخاص)، فكر في كتابة ملخص صغير (200-300 كلمة) ونشره. هذا النوع من "الأبحاث المواطنة "(Citizen Research) مهم جداً في مجال الذكاء الاصطناعي.
3. **توسيع دائرة التجارب** :جرب مواضيع أخرى. جرب لغات أخرى (إذا كنت تتحدث لغة ثانية). جرب نماذج مختلفة. قارن. كلما زادت تجاربك، زاد فهمك.
4. **ناقش التجارب مع الآخرين** .اصنع "نادي قراءة" صغير حول هذا الكتاب، واطلب من الأعضاء إجراء التجارب بأنفسهم ثم مناقشتها معاً. الحوار الجماعي حول هذه القضايا قد يكون أكثر ثراءً من القراءة الفردية.

خاتمة عامة للملحق

نبدأ من حيث انتهينا. سؤال البداية عاد إلينا متحولاً، والسطر الأخير لم يعد سطراً واحداً بل تجربة معاشة. ربما تكون هذه هي فلسفة هذا الكتاب: لا أن نعطيك إجابات، بل أن نعطيك أدوات لتبني إجاباتك الخاصة، وتختبرها بنفسك، وتنقحها، وتشاركها.

الذكاء الاصطناعي قادم، وهو هنا بالفعل. سيستمر في التطور. بعض مناظراتنا القادمة ستكون مع آلات لا نعرف أنها آلات. بعض خطاباتنا القادمة ستُكتب بآلات لا تخبرنا بذلك. بعض معتقداتنا القادمة ستتشكل بفعل خوارزميات لا نراها ولا نفهمها.

لكننا لسنا عاجزين. لأن السؤال الأهم ليس "هل تستطيع الآلة أن تقنعني؟" بل **"هل ما زلت أريد أن أكون شخصاً يقتنع بالحجج لا بالضجيج؟ يبحث عن الحقيقة لا عن الانتصار؟ يخاطر في كلامه لأنه يعني ما يقول؟"**

هذا السؤال لا تجيب عنه الآلة. ولا يكتبه خوارزم. هو سؤال الإنساني للإنساني. وطالما بقي هذا السؤال حياً، يبقى للكلام معنى. وللحياة معنى.

مراجع الملحق التفاعلي

1. **Turing, Alan M.** (1950). "Computing Machinery and Intelligence". *Mind*, 59(236), 433-460.
2. **Christian, Brian**. (2011). *The Most Human Human: What Talking with Computers Teaches Us About What It Means to Be Alive*. New York: Doubleday.
3. **Marcus, Gary, & Davis, Ernest**. (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.
4. **Mitchell, Melanie**. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.
5. **Webb, Amy**. (2019). *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*. New York: PublicAffairs.

في أحد فصول هذا الكتاب،

يروي المؤلف تجربة شخص يسأل نموذجًا لغويًا عن معنى الحب. النموذج يجيب إجابة متقنة، منظمة، شاملة، فلسفيًا صحيحة. لكن القارئ يشعر بالبرد. ليس لأن الإجابة خاطئة، بل لأنها لا تنبض.

ثم يروي الإنسان قصته: عن جده الذي كان يخبئ الحلوى في جيبه فتذوب وتلتصق بثيابه، وعن رائحة الحلوى التي كانت تخبره بقدوم جده قبل أن يراه. وهنا، يشعر القارئ بالدفء.

الفرق بين "المعلومات" و"المعنى". بين "الإقناع" و"الإلهام". بين "الكلام الصحيح" و"الكلام الحي"."

هذا الكتاب هو محاولة لاستعادة هذا الفرق في عصر طمست فيه الآلات الحدود.


DIALECTIC
διάλεκτική
RHETORIC
ρητορική τέχνη
ALGORTHM
DATA
NEURAL
NETWORK
DATA
LANGUAGE
MODEL
ETHOS
ETHOS

المحتويات

في أحد فصول هذا الكتاب، يروي المؤلف تجربة
شخص يسأل نموذجًا لغويأ عن معني الحب.
النموذج يجيب إجابة متقنة، منظمة، شاملة،
فلسفيأ صحيحة.
لكن القارئ يشعر بالبرد.
ليس لأن الإجابة خاطئة، بل لأنها لا تنبض.

ثم بروي الإنسان قصته: عن جده الذي كان
يخبئ الحلوى في جيبه فتذوب وتلتصق بثيابه،
وعن رائحة الحلوى التي كانت تخبره بقدوم جده
قبل أن يراه.
وهنا، يشعر القارئ بالدفء.

الفرق بين "المعلومات" و "المعنى".
بين "الإقناع" و "الإلهام".
بين "الكلام الصحيح" و "الكلام الحي".

هذا الكتاب هو محاولة لاستعادة هذا الفرق
في عصر طمست فيه الألات الحدود.

